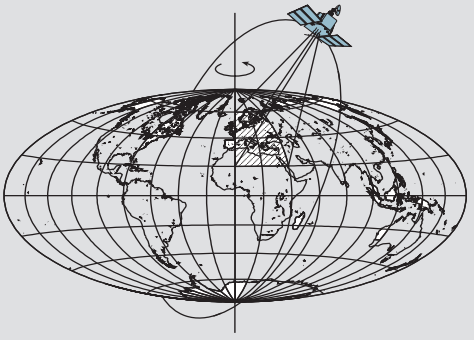
Investigations Into Green's Function as Inversion-Free Solution of the Kriging Equation, With Geodetic Applications

by

Ching-Chung Cheng



Report No. 472

Geodetic and GeoInformation Science
Department of Civil and Environmental Engineering and Geodetic Science
The Ohio State University
Columbus, Ohio 43210-1275

December 2004

INVESTIGATIONS INTO GREEN'S FUNCTION AS INVERSION-FREE SOLUTION OF THE KRIGING EQUATION, WITH GEODETIC APPLICATIONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Ching-Chung Cheng
****

The Ohio State University
2004

Dissertation Committee:

Professor Burkhard Schaffrin, Adviser

Approved by

Professor C. K. Shum

Professor Ralph von Frese

_____
Adviser

Graduate Program in Geodetic
Science and Surveying

ABSTRACT

Statistical interpolation has been proven to be a legitimate and efficient approach for data processing in the field of geodetic and geophysical sciences. Pursuing the minimization of the mean squared prediction error, the technique, known as Kriging or least-squares collocation, is able to densify, respectively filter a spatially and/or temporally referenced dataset, provided that its associated covariance model is given or estimated in advance. The involvement of the covariance matrix which to some extent reflects the physical behavior of the underlying process may, however, potentially lead to an ill-conditioned situation when the data are observed at a relatively high sampling rate.

A new perspective, interpreting the Kriging equation in the continuous sense, is therefore proposed in this research so that, instead of matrix terms, a convolution equation is set up for the Green's function where the covariance function is preserved in its analytic form. Two methods to approximate the solution of such a convolution equation are employed: One transforms the unknown Green's function into a series consisting of a linear combination of (partial) derivatives of the covariance function so that the approximation of the Green's function can be determined through a term-by-term approach; the other one manipulates the convolution equation in the spectral domain where the inversion can be treated within the space of real number.

The proposed approach has been applied to various covariance models, especially several more recently established spatial-temporal models which have attracted increasing interests for geophysical applications. Examples from geodetic science include the cases of data fusion and terrain profile monitoring; although based on simulated data, the demonstration of this innovative approach shows great potential.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Nowadays, data interpolation is among the most significant tasks for which geostatistics can provide solutions in the disciplines of geophysics and geodesy. Demands for it often arise when the sampling rate of a geo-referenced dataset is below the required sampling density, or some sites of interest are overshot by the sampling scheme. Moreover, when point-against-point comparison between two datasets is to be made, the "relocation" of data sites from one of the dataset is always mandatory. Uniquely different from other interpolation techniques, the legitimacy of statistical interpolations is based on the principle of regarding the dataset as outcome from one single stochastic experiment. As a matter of result, the so-called interpolation can be interpreted as a prediction of the next outcome, but over new sites, because the concept of repetitive experiments in geostatistics is usually generalized to include probability measures spatially and temproally. Thus, a predictor or function of already sampled data can take the role of an interpolation function. In light of this perspective, it is easy to understand that statistical interpolation, driven by the probability model, has the edge over other techniques when there is no exact physical knowledge available to model the dataset.

Among the statistical predictors, Kriging characterized by its linear form, without any doubt, is among the most efficient and practical statistical techniques. It is similar to what became known as least-squares collocation in the geodetic community, and is able to reflect the physical behavior of the underlying process to some extent. Considering only the pair-wise relations of sites (data points), Kriging predicts the signal at a new site s_0 by the linear combination of the signal at surrounding sites, using as interpolator a coefficient vector λ where each entry assigns a "measure of influence" to each data site. In the case that a constant mean of a dataset can be assumed, under the criterion of minimum mean squared prediction error, the $n\times 1$ coefficient vector λ results from the Simple Kriging equation:

$$C_{zz}\lambda = C_{z0},$$

where C_{zz} is the $n\times n$ variance-covariance matrix of the data vector, and C_{z0} is a $n\times 1$ vector sequentially recording covariances between the signal $Z(s_0)$ and each entry in the data vector $\left[Z(s_1),\cdots,Z(s_n)\right]$; n here is the sample size. However, the attempt to solve for λ, which involves the inversion of this equation, can be plagued by numerical problems when the sampling rate of data is increased to a certain extent. One can determine that every column in C_{zz}, interpreted as a discretized covariance function, would find the previous or next columns to be almost identical, but with a spatial shift from one data site to another. Therefore, high sampling rate means tiny shifts among sequential columns

and, consequently, an ill-conditioned matrix C_{zz}. As more and more cutting-edge sensor equipments become available, owing to the development of remote sensing technology such as radar imagery, it indeed shows potential demand for developing suitable algorithms to process continuous datasets (Krumm, 1987, and Borre, 2001).

To tackle this problem, a new perspective is proposed in this thesis. Conceptually, to favor the continuous properties coming from the extremely high sampling rate, the underlying random field modeling the high density dataset can be replaced by a continuous random function, and the interpolator vector λ will thus result in a continuous function $G(s)$, known as Green's function. Accordingly, the Kriging equation is generalized to a continuous case, which is the extreme form for high sampling rate datasets. Inspired by the pioneering work of Hohlfeld et al. (1993) and Oliver (1998), the Kriging equation is interpreted as a Fredholm integral equation where the unknown function $G(s)$ is to be solved for. Two approaches are shown in this thesis to solve the integral equation: The first directly transforms the equation into the representation in form of a series spanned by the partial derivatives of the covariance function; the other makes use of the Fourier transform, where the inverse is carried out in the spectral (Fourier) domain and then transformed back. Both approaches result in $G(s)$ being represented in the basis of derivatives of the covariance function, which instinctively can be associated with the sequential columns in the matrix C_{zz}.

Chapter 2 reviews the fundamental theory of random fields, as well as optimal linear predictors for data sampled from such a field. Chapter 3 investigates the theoretical background of valid covariance functions. Some important historical works are briefly introduced, and most importantly, the criteria to construct a valid covariance function are discussed in this chapter which will be referred to frequently in the following chapter. Chapter 4 addresses the approach of forming the continuous Kriging equation and presents two approaches to its solution which provide analytical representations of Green's functions by basically making use of different transforms to achieve the inverse in the continuous fashion. Following the theoretical elaboration on these approaches, some Green's function examples with respect to several well known spatial-temporal covariance functions, as well as an example of data fitting, are presented in Chapter 5. Afterwards, some promising applications in geodetic science are discussed in Chapter 6, and finally conclusions are stated in Chapter 7.

CHAPTER 2

SPATIAL-TEMPORAL DATA

To deal with applications involving uncertainties associated with spatial and temporal references, a statistical approach is often applied to model the data sets. Such a perspective regards the observed data as a (single) realized outcome of a random process defined on the referenced spatial-temporal domain, so that the mechanism which actually generated the observed data can be represented by properties of the referred random process.

2.1 Spatial-Temporal Random Fields

To stochastically model a uni-variate natural phenomenon which is spatially or temporally referenced, a group of random variables subscripted by the respective spatial location/temporal epoch are usually employed for the modeling. Conventionally, such set of random variables is denoted by

$$\{Z(s;t): s \in E,\ t \in U\} \tag{2.1.1}$$

where E is the domain where the spatial locations of all possible "sites" are accommodated, and usually but not necessarily referred to as a d-dimensional Euclidean space, denoted by $\mathbb{R}^d$. Likewise, U is the domain that contains all the temporal epochs , often characterized by the 1-dimensional time axis. In order to emphasize the casual property of time, the temporal space is further confined to the interval of $[0,\infty)$ sometimes. The lowercase letters s and t are the indicators pointing to the positions of data in the spatial and time domains, respectively. Certainly, like the definition of any random variables, any $Z(s_j;t_j): s_j \in E,\ t_j \in U$ is associated with a probability space $\Omega_j \subset \mathbb{R}^1$ with the probability measure $\Pr\{\Omega_j\} = 1$. In this thesis, however, the probability spaces of all random variables in a field are assumed to be all the same; namely, $\{\Omega_j = \mathbf{R}^1 \mid (s_j,t_j) \in E \cup U\}$.

The lowercase notation $\{z(s;t): s \in E, t \in U\}$ is employed to denote the realization from the random field $Z(s;t)$. The subscripted numbers attached to the spatial or temporal indicators s and t refer to the specific spatial location/time epoch of the data. For example, the collection of data observed at the same location, but at different time epochs can be expressed as $\{z(s_p,t_1), z(s_p,t_2),\ldots,z(s_p,t_n)\}$; similarly, the data observed at the same time, but at different locations can be written as $\{z(s_1,t_q), z(s_2,t_q),\ldots,z(s_1,t_q)\}$.

Joint Density of a Random Field

For a random field Z, the behavior of all sites at all times is governed by the associated distribution, defined by finite dimensions, which corresponds to all the data sites:

$$F\left(z(s_1;t_1),...,z(s_p;t_q)\right) \equiv \Pr\left\{Z(s_p,t_q) \le z(s_1;t_1),...,Z(s_p;t_q) \le z(s_p;t_q)\right\},$$

where

$$\frac{\partial^{p\times q} F(z(s_1;t_1),\cdots,z(s_p;t_q))}{\partial z(s_1;t_1)\cdots\partial z(s_p;t_q)} = f(z(s_1;t_1),\cdots,z(s_p;t_q)) \tag{2.1.2}$$

provided that f is a valid joint density function (Sveshnikov, 1978).

The joint distribution of a set of data sites definitely characterizes all the random properties associated with the field (process) in a complete way. Along with the analytic form of the joint density, moments of all orders (from first to infinity, if they exist) can characterize most of the unique properties of this field (process). For some distributions frequently used to model geophysical data, such as the Gaussian, the first two moments (centered) are either the most dominant, or even completely sufficient factors to determine the properties of a random field.

In compliance with the notations of most statistical textbooks, the first moment – referred to as *expectation* (also known as *probabilistic mean* or *probabilistic trend* in some occasions) – is denoted by

$$\mu(s;t) = E\{Z(s;t)\} = \int_{-\infty}^{\infty} z(s;t) f(z(s;t)) dz(s;t), \tag{2.1.3}$$

whereas the *spatial-temporal mean* is defined as

$$\mu(Z(s;t)) = \iint_{E\cup U} z(s;t) ds dt. \tag{2.1.3a}$$

Under the assumption of "ergodicity", probabilistic mean and spatial/temporal mean coincide. This is a very important issue in spatial-temporal statistics. Although the definition of moments is originally given based on the form where the density function is integrated over the probability space, in numerous practical problems, however, the only available information consists of the dataset observed over $E \cup U$, which is actually a single realization from the associated random field. In order to assess the properties of a dataset, the following definition of the *sample mean* is adopted if the ergodicity assumption holds:

$$\overline{\mu} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} z(s_i;t_j) \qquad \text{for } i \in E \text{ and } j \in U. \tag{2.1.3b}$$

As far as the centered second order moment – variance and covariance – is concerned, it indicates the degree of dispersion across the defined space and time domain. The definition of *covariance* is given by

$$C(Z(s_i;t_i), Z(s_j;t_j)) = E\{(Z(s_i;t_i) - \mu(s_i;t_i))(Z(s_j;t_j) - \mu(s_j;t_j))\} \tag{2.1.4}$$

where $s_i, s_j \in E$, $t_i, t_j \in U$. Obviously, the definition of the (co)variance relies on the fist moment, *i.e.*, the expectations μ.

Alternatively, there is another form to capture the information from the second moments, which has attracted increased attention. The *semi-variogram* is defined by

$$\gamma(Z(s_i;t_i), Z(s_j;t_j)) = \frac{1}{2} Var\{Z(s_i;t_i) - Z(s_j;t_j)\} \tag{2.1.4a}$$

and, under "ergodicity", can be expressed as

$$\gamma(Z(s_i;t_i), Z(s_j;t_j)) = E\{[Z(s_i,t_i) - Z(s_j,t_j)]^2\}. \tag{2.1.4b}$$

Here *Var* denotes the variance function, similar to (2.1.4) when *i=j*.

The semi-variogram spells out the dispersion of a process by way of the discrepancy between the signals at two data sites directly, without the direct involvement of the mean. Such property shows remarkable advantage when the mean is difficult to access.

Stationarity

In an attempt to focus on the stochastic properties of a studied process, stationarity is often assumed. This assumption depicts a simplified situation where the process has homogeneous stochastic properties across the whole space and time domains; in other words, no spatial or temporal variation is introduced by a (non-random) shift. If a random process Z is said to be *stationary*, then, for any subset of sites $\{(s_1,t_1),(s_2,t_1),\cdots,(s_m,t_n)\}$ from $E \cup U$, we have

$$F_{(s1,t1),\ldots,(sm,tn)}(Z(s_1,t_1),\ldots,Z(s_m,t_n))$$

$$= F_{(s1+s,t1+t),\ldots,(sm+s,tm+t)}(Z(s_{1+s},t_{1+t}),\ldots,Z(s_{m+s},t_{n+t})) \tag{2.1.5}$$

where (*s*, *t*) denotes a translating vector defined in $E \cup U$.

Since the probability function is often too time-consuming to handle during computations, a much simpler alternative –second-order stationarity is introduced instead. A process Z is said to be *second-order stationary* if its expectation would fulfill

$$E\{Z(s;t)\} = \mu \qquad \text{for all } s \in E \text{ and } t \in U, \tag{2.1.6}$$

and for the covariance function it holds

$$C\{Z(s_i;t_i), Z(s_j;t_j)\} = C(s_i - s_j; t_i - t_j) \qquad \text{for all } s_i, s_j \in E \text{ and } t_i, t_j \in U. \tag{2.1.7}$$

2.2 Conditional Specification and Markov Random Fields

Aside from the form of a joint distribution mentioned in (2.1.2), the spatial-temporal structure across the data sites can be specified by the conditional approach. The probability density of a single site x_i, conditioned on all the other sites x_j in the defined domain, is denoted by

$$f\langle z(x_i) | z(x_j) \rangle \text{ for all } z(x_j) \text{ where } i, j \in E \cup U \text{ but } j \neq i, \text{ and } x_j=(s_j; t_j). \tag{2.2.1}$$

In order to simplify the notation, we denote a site in the spatial-temporal domain by $x_i = (s_i; t_i)$ in the following discussion when the difference between the spatial and temporal domain is not particularly emphasized.

The conditional specification, taking the microcosm perspective, turns out to be more straightforward in stating certain statistical properties for a random process. For example, a lot of geophysical datasets, if not the majority, show very strong effect of local domination and the "influence" decreases to almost zero outside a surrounding neighborhood. Such properties can be described by the conditional specification as

$$f\langle z(x_i) | N_i \rangle = f\langle z(x_i) | Z \rangle \tag{2.2.2}$$

where Z denotes the vector collecting values from *all sample sites* in a defined domain $E \cup U$, but excluding the selected site $z(x_i)$. The symbol N_i, however, denotes a subset of Z in which only sites surrounding $z(x_i)$ are collected; it is called the *neighborhood* of site i. The identity (2.2.2) that expresses the "lack of memory" property is also known as *Makovian property*.

In the case of a stationary random field, the joint specification is intuitively supposed to be the summation of some forms of the conditional specification since those local structures are considered homogeneous everywhere. Methods to associate the joint specification with the conditional specification have been thoroughly studied by Besag (1974), and Hammersley and Clifford (1971), who introduced the concept of *cliques*, which represent various combinations of neighboring sites. By way of such

categorization as neighboring sites, a joint density can be factorized into a series of density functions conditioned on various clique sets.

2.3 The Case of a Gaussian Markov Random Field

In Section 2.2, a general perspective on the theory of Markov Random Fields was given; however, a specific example may illustrate the theory in a more practical way. The case for Gaussian data has been elaborated upon by previous researchers, and would serve as a perfect example of a typical Markov Random Field (Cressie, 1991, p.414). Let us assume a Gaussian random field that is defined on n sites, and only the pairwise relations matters, which means that all the cliques containing more than two members would be suppressed. Based on this setting, the conditional density can be written as

$$f\left\langle z(x_i)\middle| z(x_j): j \neq i\right\rangle = \left(2\pi\sigma_i^2\right)^{-1/2} \exp\left\{-\left[z(x_i) - p\left(z(x_j): j \neq i\right)\right]^2 / (2\sigma_i^{\,2})\right\} \qquad (2.3.1)$$

where f denotes the density function for site i conditioned on all the other sites. Furthermore, when the Markovian property is assumed to be effective within a q-site neighborhood nearby a centered site i , then the *conditional mean*, p, in (2.3.1) has the linear form of

$$E\left\{\left\langle z(x_i)\middle| z(x_j): j \neq i\right\rangle\right\} = \mu_i + \sum_{j=1}^{n} g_{ij}\left(z(x_j) - \mu_j\right) = \mu_i + \sum_{j=1}^{q} g_{ij} \cdot \left(z(x_j) - \mu_j\right). \qquad (2.3.2)$$

By way of the factorization theorem by Besag (1974), one can end up with the following relation

$$\log\left(f(z)/\rho\right) = \left(-\tfrac{1}{2}\right)(z-\mu)^T D^{-1}(I-G)(z-\mu) \qquad (2.3.3),$$

where ρ is a constant, $\mu = [\mu_1, \cdots, \mu_n]^T$, $G \equiv \left(g_{ij}\right)$ is an $n \times n$ matrix, and another $n \times n$ matrix is denoted by $D \equiv Diag\left(\sigma_1^2, \cdots, \sigma_n^2\right)$. Since the density function f is considered to be Gaussian, the right hand side of (2.3.3) shows the exponent of an n-dimensional Gaussian distribution, that is

$$\underset{n\times 1}{Z} \sim N\left(\mu, (I-G)^{-1} D\right) . \qquad (2.3.4)$$

2.4 Optimal Spatial-Temporal Prediction

In the effort to make an appropriate prediction, the "local structure" – density function conditioned on neighboring sites – without doubt is one of the most legitimate piece of information, unless the genuine physical model is available. If we further discard all the conditional densities, except for those defined on one-way cliques, then we confine the construction of the predictor to be necessarily linear.

Suppose the $n \times 1$ vector $z = \left(z(x_1) = z_1, \cdots, z(x_n) = z_n\right)^T$ denotes a set of observed data, and a prediction for the random variable $Z_0 \equiv Z(x_0)$ is made at the location x_0 where no observation is collected. Then a linear predictor would have the form

$$\tilde{z}_0 = \lambda^T z + \lambda_c = \sum_{i=1}^{n} \lambda_i z_i + \lambda_c \,, \tag{2.4.1}$$

where the $n \times 1$ vector $\lambda = \left(\lambda_1, \cdots, \lambda_n\right)^T$ denotes the unknown "*coefficients of influence*", and λ_c is an unknown *shift parameter*.

In order to come up with a predictor uniquely, a criterion – for instance a loss function – has to be set up so that the coefficients can be determined through its optimization. Following this doctrine, the ever popular choice of minimizing the mean squared error is adopted; it can be written as the sum of error variance and the squared bias

$$MSE\{\tilde{z}_0\} = E\left\{\left(\tilde{z}_0 - Z_0\right)^2\right\} = Var\{\tilde{z}_0 - Z_0\} + \left(E\{\tilde{z}_0 - Z_0\}\right)^2 = \min \,. \tag{2.4.2}$$

Based on this criterion, and the assumption of a given mean, the coefficient vector λ of the optimal predictor is provided by

$$C_{zz}\lambda = C_{0z} \,, \tag{2.4.3}$$

and the shift parameter λ_c by

$$\lambda_c = E\{Z_0\} - \lambda^T E\{z\}, \tag{2.4.3a}$$

where C_{zz} denotes the variance-covariance matrix of the $n \times 1$ random vector z, and C_{0z} the $n \times 1$ covariance vector between the random variable Z_0 and each variable in z. Moreover, the mean squared prediction error for the optimal predictor is given by

$$MSE\{\tilde{z}_0\} = E\left\{\left(\tilde{z}_0 - Z_0\right)^2\right\} = \sigma_{z_0}^2 - C_{0z}^T \lambda \,. \tag{2.4.4}$$

The equation (2.4.3), usually termed *Simple Kriging Equation*, has a unique solution provided that C_{zz} is nonsingular. The solution to this equation provides the Simple Kriging coefficients, and the predictor $\tilde{z}_0 = \lambda^T z + \lambda_c$ from (2.4.1) is indeed the *Best inhomogeneously Linear Predictor*.

In many applications, however, the mean of the datasets is actually unknown. One obvious approach to deal with this situation is to get "clean data" without mean drift through some preliminary treatment called "de-trending". Usually these approaches involve either the effort to cancel out the mean by multiple differencing, or to remove an estimated trend if such estimation can be made available in advance. If a constant mean μ is unknown, the *Ordinary Kriging* solution is given by

$$\tilde{\tilde{z}}(x_0) = \hat{\mu} + C_{0z}^T C_{zz}^{-1} (\underset{n\times 1}{z} - \tau\hat{\mu}) \tag{2.4.5}$$

with

$$\hat{\mu} = (\tau^T C_{zz}^{-1} \underset{n\times 1}{z}) \Big/ (\tau^T C_{zz}^{-1} \tau) \,, \tag{2.4.6}$$

where $\underset{n\times 1}{\tau} = \left[1, \cdots, 1\right]^T$.

A more general and adequate approach to handle the data with an unknown, but not necessarily constant mean consists in integrating the estimation of the mean into the Kriging estimator. To deal with the non-random mean relief, one can assume the non-random mean to be written as a finite expansion in linear form,

$$\mu_i = \sum_l \xi_l f_l(x_i), \tag{2.4.7}$$

or in matrix notation,

$$\underset{n\times 1}{\mu} = F\xi \tag{2.4.8}$$

where $\underset{l\times l}{\xi} = [\xi_1, \cdots, \xi_l]^T$,

$$\underset{n\times l}{F} = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_l(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_l(x_2) \\ \vdots & \vdots & & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_l(x_n) \end{bmatrix}, \tag{2.4.9}$$

and a singled out column vector for the site x_0:

$$f_0 = [f_1(x_0), \cdots, f_l(x_0)]^T. \tag{2.4.10}$$

At this point, let us choose the linear predictor to be unbiased, but with a homogenously linear form such that

$$\tilde{\tilde{z}}_0 = \lambda_h^T z, \tag{2.4.11}$$

with its bias vanishing:

$$E\{\tilde{\tilde{z}}_0\} = E\{\lambda_h^T z\} = E\{Z_0\}. \tag{2.4.11a}$$

By minimizing the mean squared error

$$\Phi = E\{(\tilde{\tilde{z}}_0 - Z_0)^2\} = \mathrm{var}\{\tilde{\tilde{z}}_0 - Z_0\} = \min .,$$

subject to (2.4.11-11a), the *Universal Kriging* equation

$$\begin{bmatrix} C_{zz} & F \\ F^T & \underset{l\times l}{0} \end{bmatrix} \begin{bmatrix} \lambda_h \\ \gamma \end{bmatrix} = \begin{bmatrix} C_{z0} \\ f_0 \end{bmatrix} \tag{2.4.12}$$

arises, and the accuracy of this prediction in terms of the mean squared error would be

$$MSE\{\tilde{\tilde{z}}_0\} = E\{(\tilde{\tilde{z}}_0 - Z_0)^2\} = \sigma_{z_0}^2 - \lambda_h^T C_{z0} - \gamma^T f_0 . \tag{2.4.13}$$

More details regarding the derivation of optimal predictors can be found in the comprehensive textbook by Chilès and Delfiner (1999).

CHAPTER 3

SPATIAL-TEMPORAL COVARIANCE FUNCTIONS

As discussed in Chapter 2, the structural characteristics of spatial-temporal data are determined by the associated probability distributions or, alternatively, by the associated moments. As far as most distributions are concerned, the first two moments significantly dominate the stochastic characteristics, especially when Gaussian density can be assumed where the first two moments suffice to describe all of them. Consequently, in regard of a stationary random field, the covariance function uniquely determines the coherency of a random field. In this chapter, necessary criteria to construct covariance functions for spatial-temporal data will be explored so that decent covariance model can be employed to attain the genuine properties of datasets.

3.1 Fundamental Criteria for Covariance Functions

All of the fundamental criteria required for covariance functions are, in the consequence, derived from the definition of variance, or its related properties. A covariance function, defined on a domain of any dimension, usually is not considered legitimate unless all of the criteria are fulfilled; otherwise, some problems may occur, either in the modeling of the random field, or when carrying out a prediction. At some occasions, however, those criteria, although assumed, may still have to be more or less compromised (Stein, 2003).

First, by definition, the covariance of any two random variables is a symmetric relation, $\text{cov}(Z_0, Z_h) = \text{cov}(Z_h, Z_0)$, so we have

Criterion 1: A stationary covariance function must be an even function, which means

$$C(h) = C(-h). \tag{3.1.1}$$

Then we must consider the stochastic properties of the random field; the variance at any site x_i must be larger than zero (and usually smaller than infinity), resulting in

Criterion 2: $$Var(Z_i) = C(0) > 0. \tag{3.1.2}$$

Moreover, for any two sites x_i and x_j with distance *h*, the respective variances and the covariance fulfill the Cauchy-Schwarz inequality $\left|Cov(Z_i, Z_j)\right| \le \sqrt{Var(Z_i)}\sqrt{Var(Z_j)}$; the covariance function, therefore, must fulfill

Criterion 3: $$C(0) \ge C(h). \tag{3.1.3}$$

Incidentally, as was discussed in the previous chapter, the signal Z_0 at a site x_0 of a stationary random field ought to have predictors which are linear combinations of the observed signals at all the other sites, collected in the vector $z = [z_1, \cdots, z_n]^T$; thus, once again, the variance ensures the following property:

$$0 \le Var(\tilde{z}_0 - \mu_0) = Var\left(\lambda^T (z - \mu)\right) = \tag{3.1.4}$$
$$= \lambda^T E\left\{(z - \mu)(z - \mu)^T\right\}\lambda .$$

where $\mu_0 = E\{z_0\}$ and $\mu = E\{z\}$ are the respective mean and mean vector. This leads to the *nonnegative-definiteness* of a variance-covariance matrix. As the random vector z at discrete sites is generalized to the continuous case, a covariance function is required to be positive-definite as well. Inferentially, one can check the positive-definiteness of a covariance function through

Criterion 4: A continuous function $C(h)$ is positive-definite if and only if its Fourier representation $f(\omega)$ is a positive real-valued function in the sense that

$$f(\omega) = F\{C\}(\omega) := \int e^{-i\omega h} C(h)dh \ge 0 \text{ for all } \omega, \tag{3.1.5}$$

where

$$C(h) = \frac{1}{2\pi}\int e^{i\omega h} f(\omega)d\omega = F^{-1}\{f\}(h) . \tag{3.1.6}$$

This Fourier validation was first proposed by Bochner(1955); for more geodetic references see K. P. Schwarz et al. (1990) . Notice that the Fourier transform in equation (3.1.5) will end up being a real-valued function provide that the covariance function $C(h)$ fulfills Criterion 1 (Kreyszig, 1983, p.175).

3.2 Separable Spatial-Temporal Covariance Functions

So far, all of the criteria address only the validation of covariance functions in the space domain alone. For many applications in geophysics, data are labeled by spatial locations and time spot together, which means that a covariance function ought to illustrate statistical relations over both spatial and temporal spaces. The very straightforward method to construct such a spatial-temporal covariance function involves the selection of a valid covariance function that is defined for the signals in the space $\mathbb{R}^{d+1}$, thereby devoting one dimension for time. In most cases, by following this method, the covariance function can be decomposed into the product of two functions that are defined solely on spatial parameters, or on the temporal parameter, respectively.

As far as the validation of these combined covariance functions is concerned, the product combinations also have to fulfill the criteria listed in Section 3.1. Among all the criteria, the positive-definiteness is without any doubt the most demanding one, since it requires the Fourier transform, which is not always analytically feasible. One easy instinctive method to obtain a positive-definite covariance function in space and time is to multiply positive-definite functions on either domains with each other.

Proposition 3.2.1: If $f(h)$ and $g(u)$ are both positive-definite functions, the product function, $k(h;u) = f(h)g(u)$, is also positive-definite.

Proof: Since $f(h)$ and $g(u)$ are both positive-definite, we have

$$\int f(h)e^{ih^{\mathrm{T}}\omega}dh > 0\,, \tag{3.2.1}$$

and

$$G(\tau) = \int g(u)e^{iu^{\mathrm{T}}\tau}du > 0\,. \tag{3.2.1a}$$

As a result,

$$F\{k\}(\omega;\tau) = \int\int f(h)g(u)e^{i(h^{\mathrm{T}}\omega+u^{\mathrm{T}}\tau)}dhdu = \int f(h)e^{ih^{\mathrm{T}}\omega}\left[\int g(u)e^{iu^{\mathrm{T}}\tau}du\right]dh =$$

$$= \int f(h)G(\tau)e^{ih^{\mathrm{T}}\omega}dh = G(\tau)\left[\int f(h)e^{ih^{\mathrm{T}}\omega}dh\right] > 0$$

because both (3.2.1) and (3.2.1a) are positive functions.

In light of this proposition, any available covariance function $C(h)$, $h \in \mathbb{R}^{\mathrm{d}}$, defined in a *d*-dimensional spatial domain, multiplied by a positive-definite covariance function $C(u)$ defined in the one-dimensioned time domain, can lead to an eligible spatial-temporal covariance function. On the other hand, covariance functions of this kind can always be decomposed into the product of several positive-definite covariance functions; so the name *separable covariance function* is attached to this family. For example, a Gaussian covariance function, defined in the space of *x* and *y*, joined with a Gaussian covariance functions, defined in *t*, would be

$$C(x,y,t) = \frac{1}{(abc)^{2/3}\pi}\exp\left(-\frac{x^2}{a^2}-\frac{y^2}{b^2}-\frac{t^2}{c^2}\right). \tag{3.2.2}$$

Mathematically, covariance functions from such a family are valid covariance functions; their application to real data, however, is quite limited since no interaction between space and time domains is modeled (Cressie and Huang, 1999).

On the other hand, not any linear combination of spatial covariance functions with a temporal covariance function guarantees that the newly constructed function will be a valid covariance function. The linearity of the Fourier transform would turn this into the same linear combination of two positive functions in the spectral domain. This is valid, for instance, if addition and multiplication with positive coefficients is exclusively involved in the combination; otherwise, the new one could still end up being negative in the spectral domain and, consequently, would not be a valid covariance function.

3.3 Space-Time Covariance Functions Proposed by Cressie and Huang

To overcome the drawback of separable covariance functions, Cressie and Huang (1999) proposed an alternative principle to create a new family of non-separable covariance functions $C(h;u)$.

Proposition 3.3.1: A continuous, integrable function $C(h;u)$, defined on $\mathbb{R}^d \times \mathbb{R}^1$, is a valid space-time covariance function if and only if

$$f_\omega(u) = F_C(\omega;u) = \int e^{-ih^T\omega} C(h;u)dh \tag{3.3.1}$$

is a covariance function for almost all $\omega \in \mathbb{R}^d$.

The temporal parameter u and the spatial spectral parameter vector ω are tangled together in (3.3.1) to ensure the interaction between space and time domain. However, some separable parts should still be allowed in the covariance representation, so that it can be multiplied by other positive functions in the spatial spectral parameter vectors $g(\omega)$, which action still keeps the entire covariance function positive in the spectral domain. As a consequence, a non-separable spatial-temporal covariance function $C(h; u)$ can be generated by the spatial inverse Fourier transform of

$$f_\omega(u)g(\omega) = g(\omega)\int e^{-ih^T\omega} C(h;u)dh\,. \tag{3.3.2}$$

It is not difficult to assemble some "ready-to-use" covariance functions, defined in spatial and in temporal domains, into one function which fulfills (3.3.1). However, for any arbitrary combination of functions, challenges arise when the representation (3.3.2) in the spectral domain needs to be transformed backwards to the original space-time domain. To construct closed-form examples of spatial-temporal covariance functions, Cressie and Huang (1999) make use of covariance functions and spectral density functions, given by Matérn (1960), which assure an analytical form in the original domain.

3.4 Gneiting's Theorem of Space-Time Covariance Functions

The method proposed by Cressie and Huang (1999) provides a feasible way to generate limited classes of non-separable spatial-temporal covariance functions. However, the Fourier transform in (3.3.2) still cannot be avoided, which often turns out to be a quite demanding job. To further explore a method covering a more generalized class of covariance functions, Gneiting (2002) attempted to work on the alternative criteria in the original domain, rather than in the spectral domain.

Proposition 3.4.1: If $\varphi(t)$, for $t \geq 0$ is a completely monotone function, and $\psi(t)$, for $t > 0$ is a positive function with a completely monotonic derivative, then

$$C(h;u) = \frac{\sigma^2}{\psi\left(|u|^2\right)^{d/2}} \varphi\left(\frac{\|h\|^2}{\psi\left(|u|^2\right)}\right) \tag{3.4.1}$$

is a spatial-temporal covariance function.

Definition: An analytic function $f(t)$ in $\mathbb{R}^d$ is said to be *completely monotone* if it possesses derivatives $f^{(n)}$ of all orders $n \in \mathbb{N}$ that fulfill the inequalities

$$(-1)^n f^{(n)}(t) \geq 0\,. \tag{3.4.2}$$

From the form of (3.4.1), one can easily conclude that any covariance function of this form will be an even function, that is, $C(-h;-u) = C(h;u)$. The property ensures that all covariance functions of the Gneiting form will have a real-valued Fourier transform in the spectral domain (Kreysig, 1983, p.175).

3.5 Smoothness of Covariance Functions

Aside from the propositions mentioned in the previous sections of this chapter, Stein (1999) further argues that any lack of smoothness of a spatial-temporal covariance function around the origin would cause disturbances if the locations of the observations involved in the linear predictor had any small deviations. In other words, the linear predictor would be fairly unstable when small scale errors are introduced into the location of observations. In consideration of this point, a covariance function is required to possess some smoothness, or more specifically, differentiability, at least around the area near the origin; otherwise, some numerical problems may occur in practical applications.

Once again, such a property can better be studied in the spectral domain. To enforce the requirement of this differentiability on covariance functions, Stein (2003) showed that, if the derivatives of a Fourier transformed density have certain moments, then the corresponding covariance function possesses derivatives away from the origin. Based on this theorem, a rich class of Fourier transformed densities (probability density) has been proposed to have the form of

$$f(\omega,\tau) = \left\{ c_1 (a_1^2 + |\omega|^2)^{\alpha_1} + c_2 \left(a_2^2 + |\tau|^2 \right)^{\alpha_2} \right\}^{-\nu}, \qquad (3.5.1)$$

with parameters c_1, c_2, a_1, a_2 and properly chosen exponents α_1, α_2, ν, where $\alpha_1^2 + \alpha_2^2$ is positive and

$$\frac{d_1}{\alpha_1 \nu} + \frac{d_2}{\alpha_2 \nu} < 2 \qquad \text{for } d_1 \text{ as spatial and } d_2 \text{ as temporal dimension} \qquad (3.5.2)$$

Analytically, the inverse Fourier transform of such a density has turned out to be quite difficult to compute; so far, closed forms seem only possible for some special cases involving choices of α_1, α_2, and ν, as shown by Stein (2003). However, from the perspective in Chapter 4 below, such representation in the spectral domain is still very appreciable, considering the inverse problem in constructing Green's function. As a matter of fact, it is easier to invert a function in the spectral domain than in the original space-time domain provided that an analytical form is available. As far as the representation in the original space-time domain is concerned, there is almost never a lack of numerical tools to overcome such a problem in its implementation.

3.6 Convolution Operators

In order to assure the positive-definiteness, new covariance functions can be generated by the convolution of various functions. The first example is the approach of *auto-convolution*.

Proposition 3.6.1: For any even and continuous function $q(x): \mathbb{R}^d \to \mathbb{R}^1$, $x \in \mathbb{R}^d$, with the Fourier transform $Q(\omega): \mathbb{R}^d \to \mathbb{R}^1$, $\omega \in \mathbb{R}^d$, the self-convolution $W(x) = q(x) * q(x)$ is a positive-definite function.

Proof: The spectral representation of $W(x)$ is real-valued with

$$F\{W\}(\omega) = F\{q\}(\omega) \cdot F\{q\}(\omega) = [Q(\omega)]^2,$$

which

h is a positive function. So the composite function $W(x)$ is a positive-definite function.

Another possible way to come up with a new positive-definite function is through the convolution of two other positive-definite functions in accordance with Proposition 3.2.2. A good example is when the new function $\rho(x): \mathbb{R}^d \to \mathbb{R}^1$ is unknown and involved in a convolution equation,

$$\rho(x) * K(x) = \psi(x), \tag{3.6.1}$$

where $K(x)$ and $\psi(x)$ are both positive-definite mappings of type $\mathbb{R}^d \to \mathbb{R}^1$. After taking the Fourier transform on both sides, the solution can be written as

$$\rho(x) = F^{-1}\left\{\frac{1}{F\{K\}(\omega)}\right\}(x) * \psi(x), \tag{3.6.2}$$

which is the convolution of two positive-definite functions; so $\rho(x)$ is positive-definite.

A typical example is the Kriging equation, where the solution of Green's function will be represented by the convolution of a covariance function with other positive-definite terms. So the Green's function, as a byproduct, can potentially serve as a new covariance model. More details will be discussed in Section 4.8 below.

However, the other criteria discussed previously are also mandatory, besides the positive-definiteness, for a new function to qualify as a covariance model. Generally, to make the new function fulfill those criteria, the "ingredient" functions should follow them because of the properties of convolution operation.

Proposition 3.6.2: Suppose $q(x)$ is an even and squared-integrable function, then the new function $W(x)$ created by self convolution of $q(x)$, $W(x) = q(x) * q(x)$, is also an even function.

Proof: Since we know $q(-x) = q(x)$, so the self-convolution exists and reads

$$W(-x) = \int_{-\infty}^{\infty} q(-x - x_u) q(x_u) dx_u = \int_{-\infty}^{\infty} q(x - (-x_u)) q(-x_u) d(-x_u) =$$

$$= \int_{-\infty}^{\infty} q(x - x_v) q(x_v) dx_v = q(x) * q(x) = W(x),$$

thus showing that $W(x)$ is also an even function. This proposition can be easily generalized into the case of $x \in \mathbb{R}^d$ by

$$W_{x_1\cdots x_c}(-x_1,\cdots,-x_d)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}q(-x_1-x_{u1},\cdots,-x_{\mathrm{d}}-x_{u\mathrm{d}})q(x_{u1},\cdots,x_{u\mathrm{d}})dx_1\cdots dx_u=$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}q(x_1-(-x_{u1}),\cdots,x_{\mathrm{d}}-(-x_{u\mathrm{d}}))q(-x_{u1},\cdots,-x_{u\mathrm{d}})d(-x_1)\cdots d(-x_{u\mathrm{d}})=W_{x_1\cdots x_d}(x_1\cdots x_{\mathrm{d}}).$$

Proposition 3.6.3: Suppose that $g(x)$ is an even and continuous function with $q(0)>0$, then the self-convolution function $W(x)$ also implies that $W(0)>0$.

Proof: $W(0)=\int_{-\infty}^{\infty}q(-x_u)q(x_u)dx_u=\int_{-\infty}^{\infty}[q(x_u)]^2dx_u$.

Now, $[q(x)]^2$ is an all non-negative function, and $q(x)>\delta>0$ for some $|x|<\varepsilon$ due to continuity; this implies $W(0)>0$. Once again, this proposition can be generalized to the case of $x\in\mathbb{R}^{\mathrm{d}}$ where the integral

$$W_{x1\cdots xd}(0,\cdots,0)=\int\cdots\int[q(x_{u1},\cdots,x_{u\mathrm{d}})]^2dx_{u1}\cdots dx_{u\mathrm{d}}$$

would be larger than 0 as well.

Proposition 3.6.4: Suppose $q(x)$ is even and continuous with $q(0)\geq q(h)$ for all $h\in\mathbb{R}$, then also $W(0)\geq W(h)$ for all $h\in\mathbb{R}$.

Proof: $W(h)=\int q(h-x_u)q(x_u)dx_u=\int q(x_u-h)q(x_u)dx_u$

Let $f(x_u)$ denote $q(x_u)$ and $g_h(x_u)$ denote $q(x_u-h)$; then by Cauchy-Schwartz inequality, we have

$$[W(h)]^2=\left[\int f(x_u)g_h(x_u)dx_u\right]^2\leq\int|f(x_u)|^2dx_u\cdot\int|g_h(x_u)|^2dx_u. \qquad (3.6.3)$$

The left hand side of (3.6.3) reaches its maximum when $f(x_u)=g_h(x_u)$, that means when $h=0$. So, we can conclude that $W(0)\geq W(h)$.

CHAPTER 4

THE REPRESENTATION OF GREEN'S FUNCTION

In this chapter, the numerical properties regarding the linear predictor associated with a random field are further explored. Let the random field be denoted by

$$\{Z(x) = Z(s,t) = Z(s_1,...,s_d,t) : x \in D\}, \qquad D \subseteq \mathbb{R}^{d+1}.$$

In order to concentrate on the random properties, let us assume the random process Z to be second order stationary throughout the discussions in this chapter, unless otherwise specified. Furthermore, there is no need to set up a boundary for the domain D, which means that it can stretch from minus infinity to infinity in each dimension.

In the case when the mean is given, the best linear (unbiased) predictor of Z_0, also known as the predicted conditional mean, is shown in (2.4.1) in conjunction with (2.4.3-3a) as

$$\tilde{z}_0 = \mu_0 + C_{zo}^T C_{zz}^{-1}(\mathrm{z}-\mu) = \mu_0 + \lambda^{\mathrm{T}}(\mathrm{z}-\mu), \qquad \text{where } \mu_0=\mathrm{E}\{Z_0\}, \mu=\mathrm{E}\{z\}.$$

However, the wonderful theoretical structure above will sometimes cause problems when performing the numerical computations, as each column in C_{zz} is a vector depending on displacements in the domain D relative to the next column. If the displacement is very small, the matrix will show the problem of ill-conditioning and, consequently, the numerical computation of an inverse is not feasible.

4.1 Green's Function and a Continuous Form of Kriging Coefficients

Despite the fact that the discrete covariance matrices/vectors are often employed in the models, the sampling of a random field can actually be further generalized to continuously cover the spatial and temporal domains. This portrays a perspective that the originally finite, discretized data vector is instead selecting continuous samples from a field whose domain is a continuum. Especially when the sampling rate is quite high, it would be more appropriate to replace the vector representation with a function representation.

Therefore, the linear predictor (2.4.1) after applying (2.4.3a) is modified as

$$\tilde{z}_0 = \mu_0 + \int_{\mathrm{D}\cup\mathrm{T}} \lambda(s_0,t_0,s,t)[\mathrm{z}(s,t) - \mu(s,t)]\mathrm{d}s\mathrm{d}t\,. \tag{4.1.1}$$

Such a perspective broadens our view to look at the Kriging coefficient vector as continuous function $\lambda(s_0,t_0,s,t)$ in both (s,t) and (s_0, t_0), which is, in fact, associated with the Green's function to represent the solution of certain boundary value problems. Suppose we have the stochastic boundary values

$$z(v) = Z(v) + e(v) \qquad \text{for } v \in \partial\Omega \tag{4.1.2a}$$

collected and wish to determine

$$Z(v_0) = Z_0 \qquad \text{for any } v_0 \in \Omega\,. \tag{4.1.2b}$$

The general solution of this boundary value problem may formally be represented as:

$$Z(v_0) = \mu_0 + \int_{\partial\Omega} G(v_0, v)[Z(v) - \mu(v)]dv\,, \tag{4.1.3}$$

if the respective Green's function $G(v_0, v)$ is readily available. Even then $Z(v)$ needs to be replaced by the observed $z(v)$ which leads to the *prediction* of Z_0 through

$$\tilde{z}_0 = \mu_0 + \int_{\partial\Omega} G(v_0, v)[z(v) - \mu(v)]dv = Z_0 + \int_{\partial\Omega} G(v_0, v)e(v)dv\,. \tag{4.1.4}$$

By comparison with (4.1.1), the Kriging coefficients turn out to be (descretized) values of the appropriate Green's function.

In many cases, especially when data come in high resolution, the construction of the linear predictor form is plagued by numerical problems in forming the matrix inverse. But, for the case that a Green's function exists, the following sections discuss approaches to overcome the numerical difficulties in pursuit of the approximated Green's function.

4.2 Densification of Kriging Coefficients

An intuitive approach to pursue Green's function is to densify the discrete Kriging coefficients vector through some interpolation techniques. Figure 4.1 superimposes the Kriging coefficients calculated from three sub-datasets that were derived from an original dataset with sampling rate h by re-sampling into these three sub-datasets (denoted by red cross, blue cross and blue circle, respectively) with lower sampling rate $3\times h$.

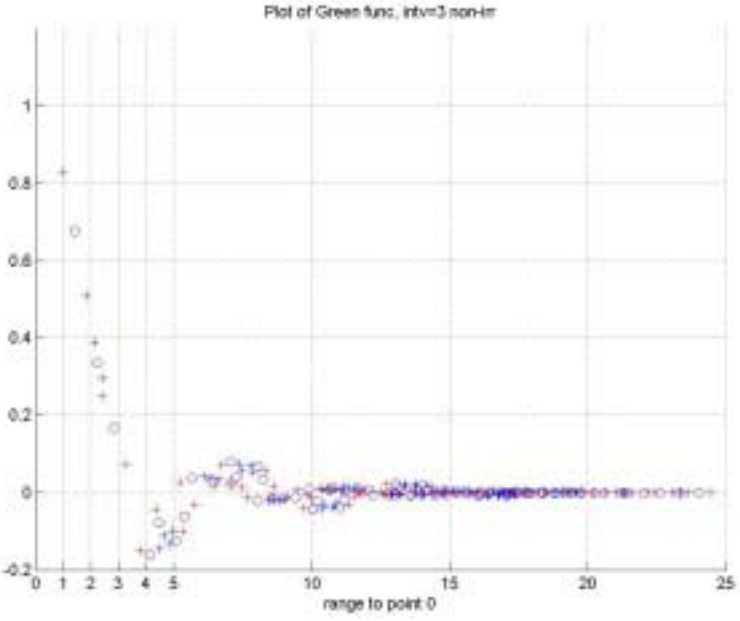


Figure 4.1: The superimposition of three lower resolution datasets

It shows that the low resolution samplings still result in Kriging coefficients that are approximately consistent with these coefficients calculated from the original high-resolution dataset. Therefore, to make up for any deficiency in resolution, one can interpolate the Kriging coefficients to fulfill the requirements. Since we have

$$\underset{n\times 1}{G}(s_0,t_0,s,t)\approx C_{zz}^{-1}C_{zo}=\underset{n\times 1}{\lambda}(s_0,t_0,s,t) \qquad \text{for fixed } (s_0,\, t_0), \tag{4.2.1}$$

some interpolation function $p(\lambda)$ goes through each component in the vector λ, corresponding to a site (s,t) within the covariance matrix/vector. We could thus obtain an approximation of the associated Green's function, by identifying $p(\lambda)\cong G(s_0,t_0;s,t)$, for arbitrary (s_0,t_0) and (s,t).

4.3 Continuous Representation for the Kriging Equation

In accordance with the perspective proposed in Section 4.1, a continuous variation to the Kriging equation (2.4.3) is given by

$$C\{z_i,Z_0\}=C_0(x_i,x_0)=\frac{1}{|D|}\int_D C(x_i,x_j)G(x_j,x_0)dx_j \quad \text{for } x_i,x_j,x_0\in D=\mathbb{R}^{\mathrm{d}+1} \tag{4.3.1a}$$

$$\approx\sum_D C(x_p,x_q)G(x_q,x_0) \qquad \text{for } x_p,x_q,x_0\in D=\mathbb{R}^{\mathrm{d}+1}, \tag{4.3.1b}$$

where $C\{z_p,z_q\}=C(x_p,x_q)$.

The mechanism of integration in (4.3.1a) corresponds to that of the multiplication of vectors/matrices. Suppose we first start with the discretized random field Z, in which, all the n data sites are sequentially labeled by a one-dimensional index $x_1,\cdots,x_n$, while x_0 stands for the location at which the prediction is to be made. The discrete Kriging equation would take the form of (4.3.1b). To illustrate the mechanism of vector multiplication in the Kriging equation, Figure 4.2 shows a segment cut from a very large Kriging equation of matrix/vector form. Here every column (or row) in the matrix represents the covariance function relative to a "center" site, which varies from x_{p-1} through x_p, to x_{p+1} (or from x_p to x_{p+3}).

$$\begin{bmatrix} \vdots \\ C_0(x_p,x_0) \\ C_0(x_{p+1},x_0) \\ C_0(x_{p+2},x_0) \\ C_0(x_{p+3},x_0) \\ \vdots \end{bmatrix} = \begin{bmatrix} & & \vdots & & \\ \cdots & C(x_p,x_{p-1}) & C(x_p,x_p) & C(x_p,x_{p+1}) & \cdots \\ \cdots & C(x_{p+1},x_{p-1}) & C(x_{p+1},x_p) & C(x_{p+1},x_{p+1}) & \cdots \\ \cdots & C(x_{p+2},x_{p-1}) & C(x_{p+2},x_p) & C(x_{p+2},x_{p+1}) & \cdots \\ \cdots & C(x_{p+3},x_{p-1}) & C(x_{p+3},x_p) & C(x_{p+3},x_{p+1}) & \cdots \\ & & \vdots & & \end{bmatrix} \begin{bmatrix} \vdots \\ G(x_{p-1},x_0) \\ G(x_p,x_0) \\ G(x_{p+1},x_0) \\ \\ \vdots \end{bmatrix}$$

Figure 4.2: The segment of matrix multiplication in the Kriging equation

Such multiplication mechanism finds its analog in the integration in equation (4.3.1a). In the continuous form, the integration variable x_j varies over the integration domain while the column index moves through $\{\ldots,p-1,p,p+1,\cdots\}$ in the covariance matrix, and, in the mean time, the first index of the Green's function varies through all

the data sites. In such a perspective, the integration over the parameter x_j inherits the algebra of multiplication between a matrix and a vector.

We recall from Section 2.1 that the site location increment is the essential variable for covariance functions (and for Green's functions as well); so, instead of the location of two sites, one may replace the function variables (x_i, x_j) with the incremental vector $(x_i - x_j)$ in case of stationarity, leading to

$$C\{z_i, z_j\} = C(x_i, x_j) = C(x_i - x_j).$$

Consequently, the continuous Kriging equation would turn to

$$C\{z_i, Z_0\} = C_0(x_i - x_0) = \frac{1}{|D|}\int_D C(x_i - x_j)G(x_j - x_0)dx_j \,. \qquad (4.3.2a)$$

As stationarity is assumed, the covariance function would be invariant to a parallel shift to the location anywhere in the field. Consequently, without lost of generality, one may set up the coordinate origin of D at x_0, which means that all variables in equation (4.3.2a) are displaced by $-x_0$, so the notation in equation (4.3.2a) would turn to

$$C_0([x_i - x_0] - [x_0 - x_0]) =$$

$$= \frac{1}{|D|}\int_D C([x_i - x_0] - [x_j - x_0])G([x_j - x_0] - [x_0 - x_0])d(x_j - x_0) \Rightarrow$$

$$\Rightarrow C([x_i - x_0]) = \frac{1}{|D|}\int_D C([x_i - x_0] - [x_j - x_0])G([x_j - x_0])d(x_j - x_0)\,. \qquad (4.3.2b)$$

Now let $x_i - x_0$ be replaced with x, and $x_j - x_0$ with x_u. Although the new notations x and x_u are in fact incremental vectors, yet they can also be interpreted as the site locations with respect to a coordinate system defined at x_0. By way of this notation, replacements of the covariance function and Green's function in (4.3.2b) can be written as

$$C(x_i - x_0) = C(x_0 + x, x_0) = C_0(x),$$

and $$G(x_j - x_0) = G(x_0 + x_u, x_0) = G(x_u).$$

The site locations x and x_u, in the coordinate system centered at x_0, become the sole parameters for a covariance function or Green's function. However, the subscript in $C_0(x)$ denoting the origin of the coordinate system cannot be erased, although the stationarity grants the invariance property; it is necessary to indicate that the process value Z_0 is to be predicted on the basis of the noisy sample value in z. Figure 4.3 shows a covariance function $C(x)$ centered at the origin $[0,0]^T$, while the covariance function $C(x - x_u)$ is identical to the previous one, but now centered at x_u.

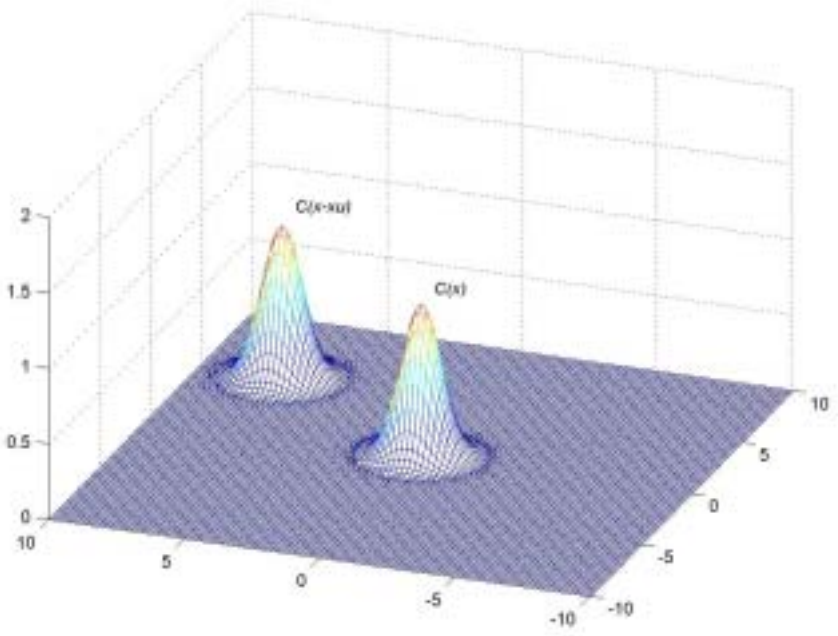


Figure 4.3: The shift of covariance functions

As a consequence, equation (4.3.2b), when introduced with the convolution operator $*$, takes the form of

$$\boxed{C_0(x) = \int_D C(x - x_u)G(x_u)dx_u = C(x) * G(x) \cong \sum_{x_i \in D} C(x - x_i)G(x_i)}, \qquad (4.3.2c)$$

provided that $C(x)$ and $G(x)$ are well normalized over D. Apparently the term $x - x_u$, instead of a displacement between sites, actually represents a shifting vector moving the covariance function $C(x)$ through the entire domain D. In such a perspective, the integrand in equation (4.3.2c) can be interpreted as the superposition of $C(x)$, at various locations of the domain D, with weighting values given by $G(x)$. Figure 4.4 shows the superimposition of a covariance vector rewritten from the segment shown in Figure 4.2. On the right hand side, two covariance vectors centered at x_{p-1} and x_p, are multiplied by $G(x_{p-1})$ and $G(x_p)$, respectively; together with the covariance vectors centered elsewhere in D and their corresponding Green's function values, their linear combination turns into the covariance vector centered at x_0.

$$\begin{bmatrix} \vdots \\ C_0(x_p\ \ ) \\ C_0(x_{p+1}) \\ C_0(x_{p+2}) \\ C_0(x_{p+3}) \\ \vdots \end{bmatrix} = \cdots + \begin{bmatrix} \vdots \\ C(x_p - x_{p-1}) \\ C(x_{p+1} - x_{p-1}) \\ C(x_{p+2} - x_{p-1}) \\ C(x_{p+3} - x_{p-1}) \\ \vdots \end{bmatrix} G(x_{p-1}) + \begin{bmatrix} \vdots \\ C(x_p - x_p) \\ C(x_{p+1} - x_p) \\ C(x_{p+2} - x_p) \\ C(x_{p+3} - x_p) \\ \vdots \end{bmatrix} G(x_p) + \cdots$$

Figure 4.4: Superimposition of covariance vectors from different locations

For the case when the dimension of $x \equiv [x_1, \cdots, x_d]^T \in D$ is d, it is not difficult to generalize the forms from the case of $\mathbb{R}^1$ because the convolution is, in fact, composed of layers of integrals with respect to each dimension:

$$C_0(x_1, \cdots, x_{\mathrm{d}}) = \int_D C(x - x_u) G(x_u) dx_u =$$

$$= \int_{D_1} \cdots \int_{D_{\mathrm{d}}} C(x_1 - x_{u1}, \cdots, x_{\mathrm{d}} - x_{u\mathrm{d}}) G(x_{u1}, \cdots, x_{u\mathrm{d}}) dx_{u1} \cdots dx_{u\mathrm{d}}$$

By way of the above interpretation, equation (4.3.2c), in the form of a Fredholm integral equation of the first kind, is now the continuous counterpart for the Kriging equation (2.4.3) where, through the convolution operator (including the integral and the kernel $C(\cdot)$), the function $G(x): \mathbb{R}^{\mathrm{d}} \to \mathbb{R}^1$ is mapped into $C_0(x): \mathbb{R}^{\mathrm{d}} \to \mathbb{R}^1$, and $G(x)$ is the unknown function. The solution of equation (4.3.2c) is supposed to have the form of

$$G(x) = R(x) * C_0(x) \tag{4.3.2d}$$

for some function R(x): $\mathbb{R}^{\mathrm{d}} \to \mathbb{R}^1$.

Equation (4.3.2d) can be compared with the discrete form as shown in Figure 4.5, where $R(x)$ stands for elements in the inverse covariance matrix C_{zz}^{-1}. Like the case of equation (4.3.1a-b), each column/row represents the function $R(x)$, evaluated at various locations.

$$\begin{bmatrix} \vdots \\ G(x_p\ \ ) \\ G(x_{p+1}) \\ G(x_{p+2}) \\ G(x_{p+3}) \\ \vdots \end{bmatrix} = \begin{bmatrix} & & \vdots & & \\ \cdots & R(x_p - x_{p-1}) & R(x_p - x_p) & R(x_p - x_{p+1}) & \cdots \\ \cdots & R(x_{p+1} - x_{p-1}) & R(x_{p+1} - x_p) & R(x_{p+1} - x_{p+1}) & \cdots \\ \cdots & R(x_{p+2} - x_{p-1}) & R(x_{p+2} - x_p) & R(x_{p+2} - x_{p+1}) & \cdots \\ \cdots & R(x_{p+3} - x_{p-1}) & R(x_{p+3} - x_p) & R(x_{p+3} - x_{p+1}) & \cdots \\ & & \vdots & & \end{bmatrix} \begin{bmatrix} \vdots \\ C_0(x_{p-1}) \\ C_0(x_p\ \ ) \\ C_0(x_{p+1}) \\ \\ \vdots \end{bmatrix}$$

Figure 4.5: Green's function discretely approximated by a matrix multiplication

In the case when the mean function is unknown, the Universal Kriging solution, as shown in equation (2.4.12), can have the continuous expressions

$$C(x) * G(x) = C_0(x) - \sum_{i=1}^{l} \gamma_i \cdot f_i(x) \tag{4.3.3a}$$

$$\int_D b_i(x) G(x) dx = f_i(x_0) \qquad \text{for } i = 1, \cdots, l; \tag{4.3.3b}$$

where $f_i(x)$ are the basis functions that span the mean function, and γ_i are the constant coefficients (scalar) to be determined.

Before proceeding to solve these integral equations, some properties of the convolution operator are worth being reviewed once again. Let f, g, and h be arbitrary functions and α a constant, then the convolution operator satisfies:

1. commutative law: $f * g = g * f$ (4.3.4a)
2. associative law: $f * (g * h) = (f * g) * h$ (4.3.4b)
3. distributive law: $f * (g + h) = (f * g) + (f * h)$ (4.3.4c)
4. $\alpha(f * g) = (\alpha f) * g = f * (\alpha g)$ (4.3.4d)
5. $\frac{d}{dx}(f(x) * g(x)) = \frac{df(x)}{dx} * g(x) = f(x) * \frac{dg(x)}{dx}$ (4.3.4e)

The proof of above properties can be found in many textbooks, including Bracewell, (1965, p.118).

4.4 Solving the Fredholm Integral Equation in $\mathbb{R}^1$

As discussed in Section 4.3, the simple Simple Kriging equation is interpreted by a convolution equation (4.3.2c), which maps the Green's function $G(x)$ into the covariance function $C_0(x)$. Under certain conditions the unknown Green's function may be determined by a numerical approach following Hohlfeld et al. (1993) where the Green's function as well as the mapping kernel $C(x,x_u)$ – expanded into the a series of pseudo-function that span the *space of hyperdistribution* and involved the *Dirac delta function*. Such representation grants a way to substitute the inversion in equation (4.3.2c) with convolution algebra so that the solution of Green's function can be approximated in terms of pseudo-derivatives of the covariance function. In this section, we will introduce the essential theorem in $\mathbb{R}^1$ because of its simplicity; it will then be generalized to $\mathbb{R}^d$ in the next section. To begin with the approach proposed by Hohfeld et al. (1993), a differential operator is introduced by

$$L_n(\nabla) = \alpha_0 + \nabla_{\alpha_1} + \cdots + \nabla^{n-1}_{\alpha_{n-1}} + \nabla^{n}_{\alpha_n}. \tag{4.4.1}$$

In case of univariate functions in x, the notation $\nabla^{p}_{a_p}$ stands for the p^{th}-order derivative with coefficient $\alpha_p(x)$; hence the operator $L(\nabla)$ can also be expressed by

$$L_n(\nabla) = \alpha_0(x) + \alpha_1(x)\frac{\partial}{\partial x} + \cdots + \alpha_n(x)\frac{\partial^n}{\partial x^n} \tag{4.4.2}$$

when applied to functions in x. Meanwhile, a second operator is notated by

$$L_n^{-1}(\nabla) = \beta_0 + \nabla_{\beta_1} + \cdots + \nabla^{n}_{\beta_n}, \tag{4.4.3}$$

which, in the univariate case, translates into

$$L_n^{-1}(\nabla) = \beta_0(x) + \beta_1(x)\frac{\partial}{\partial x} + \cdots + \beta_n(x)\frac{\partial^n}{\partial x^n}. \tag{4.4.4}$$

It is meant to be the inverse operator to (4.4.2) for suitable choices of the function $\beta_p(x)$. The Green's function in (4.3.2d), by identifying the inverse convolution operator $[R(x)*]$ with (4.4.4), is expressed as

$$G(x) = L_{\infty}^{-1}(\nabla)\big(C_0(x)\big) = L_{-\infty}^{-1}(\nabla)\big(\delta(x) * C_0(x)\big) = L_{-\infty}^{-1}(\nabla)\big(\delta(x)\big) * C_0(x) = R(x) * C_0(x)\,, \tag{4.4.5}$$

where δ(x) stands for the Dirac delta function. Meanwhile, the original convolution operator $[C(x)*]$ takes, in accordance with (4.3.4c), the form

$$C(x) * \rho(x) = L_{\infty}(\nabla)\big(\rho(x)\big) = L_{\infty}(\nabla)\big(\delta(x) * \rho(x)\big) = L_{\infty}(\nabla)\big(\delta(x) * \rho(x)\big), \tag{4.4.6}$$

where $\rho(x)$ is any arbitrary function. So, the equation (4.3.2c) now has the form

$$C_0(x) = C(x) * G(x) = L_{\infty}(\nabla)\big(\delta(x)\big) * L_{\infty}^{-1}(\nabla)\big(C_0(x)\big) =$$

$$= L_{\infty}(\nabla)\delta(x) * \big\{L_{\infty}^{-1}(\nabla)\delta(x) * C_0(x)\big\} = \big\{L_{\infty}(\nabla)\delta(x) * L_{\infty}^{-1}(\nabla)\delta(x)\big\} * C_0(x)\,, \tag{4.4.7}$$

following rule (4.3.4b), and results in the identity

$$L_{\infty}(\nabla)\big(\delta(x)\big) * L_{\infty}^{-1}(\nabla)\big(\delta(x)\big) = \delta(x)\,. \tag{4.4.8}$$

Since $C_0(x)$ is quite arbitrary. In finite approximation, this formula reads

$$\delta(x) \cong L_n(\nabla)\big(\delta(x)\big) * L_n^{-1}(\nabla)\big(\delta(x)\big) = \left(\sum_{q=0}^{n} \nabla_{\alpha}^{q}\delta(x)\right) * \left(\sum_{p=0}^{n} \nabla_{\beta}^{p}\delta(x)\right). \tag{4.4.8a}$$

As the Green's function (4.4.5) is unknown, the variable coefficients $\beta_0(x), \cdots, \beta_n(x)$ in the representation (4.4.4) are left to determine. On the other hand, since the coefficients $\alpha_0(x), \cdots, \alpha_n(x)$ in (4.4.2) are possible to be calculated by expanding the covariance function starting from the convolution with a Dirac delta function δ(x), namely

$$C(x) = \delta(x) * C(x) = \int_D \delta(x - x_u)C(x_u)dx_u \approx \sum_{x_i \in D} \delta(x - x_u)C(x_u).$$

Then, the delta function $\delta(x - x_u)$ is expanded into an analog of Taylor's series, so that the covariance function $C_0(x)$ in equation (4.3.2c) would translate into a representation in a new basis formed by the pseudo-derivatives of the Dirac delta function, namely

$$C_0(x) = C(x) * G(x) = \left\{\int_D \left(\delta(x) - \delta'(x)x_u + \frac{\delta''(x)}{2!}x_u^2 - \frac{\delta'''(x)}{3!}x_u^2 + \cdots\right)C(x_u)dx_u\right\} * G(x) =$$

$$= \left\{\delta(x)\left(\int_D C(x_u)dx_u\right) - \delta'(x)\left(\int_D x_u C(x_u)dx_u\right) + \frac{\delta''(x)}{2!}\left(\int_D x_u^2 C(x_u)dx_u\right) - \cdots\right\} * G(x) =$$

$$\cong \left\{\delta(x)\sum_{x_i \in D} C(x_i) - \frac{\delta'(x)}{1!}\sum_{x_i \in D} x_i C(x_i) + \cdots \pm \frac{\delta^{(n)}(x)}{n!}\sum_{x_i \in D} x_i^n C(x_i)\right\} * G(x) =$$

$$= \left\{\alpha_0\delta(x) + \alpha_1\delta'(x) + \alpha_2\delta''(x) + \cdots + \alpha_n\delta^{(n)}(x)\right\} * G(x)\,. \qquad (4.4.9)$$

One should keep in mind that this representation is only possible if it is placed within the convolution integrals, for instance in the sense

$$\alpha_0\delta(x) * G(x) = \alpha_0 \int_D \delta(x - x_u)G(x_u)dx_u = \alpha_0 G(x)\,,$$

$$\alpha_1\delta'(x) * G(x) = \alpha_1 \int_D \delta'(x - x_u)G(x_u) =$$

$$= \alpha_1\left\{\left[\delta(x - x_u)G(x_u)\right]_D - \int \delta(x - x_u)G'(x_u)dx_u\right\} = -\alpha_1{}' G(x)\,, \text{ etc.}$$

Incidentally, the coefficients in the vector $\alpha = \left[\alpha_0, \alpha_1, \cdots, \alpha_n\right]^T$ are associated with the *moment generator* of the covariance function as follows:

$$\alpha_j = \frac{(-1)^j}{j!}\int_D x^j C(x)dx \cong \frac{(-1)^j}{j!}\sum_{x_i \in D}(x_i)^j C(x_i)\,. \qquad (4.4.10)$$

After the coefficients α_i of the covariance representation are calculated by the moment generator, the equation (4.3.9) will be spread out, using (4.4.4-5), and then aggregated into various terms. By the Bochner-Martin algebra (Bochner and Martin, 1948), based on (4.4.8), this would lead to the identity

$$\alpha_0\beta_0\delta(x) +$$
$$+ \left(\alpha_0\beta_1 + \alpha_1\beta_0\right)\delta'(x) +$$
$$+ \left(\alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0\right)\delta''(x) +$$
$$+ \left(\alpha_0\beta_3 + \alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_0\right)\delta'''(x) + \cdots = \delta(x)$$

where the coefficients β_j are still unknown. By comparing both sides term by term, the following equations arise:

$$\begin{aligned} 1 &= \alpha_0\beta_0\,, \\ 0 &= \alpha_0\beta_1 + \alpha_1\beta_0\,, \\ 0 &= \alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0\,, \\ 0 &= \alpha_0\beta_3 + \alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_0\,, \quad \text{etc} \end{aligned} \qquad (4.4.11)$$

The solution of integral equation (4.3.2c), finally surfaces once the coefficients $\beta = \left[\beta_0, \cdots, \beta_n\right]^T$ are solved through (4.4.11) sequentially, provided that the vector $\alpha = [\alpha_0, \cdots, \alpha_n]^T$ had been obtained through equation (4.4.10).

As far as the solution of Universal Kriging is concerned, one can regard the right hand side of equation (4.3.3a) as a compound function $\phi(x; \gamma_1, \cdots, \gamma_l)$, so it takes the form

$$C(x) * G(x) = C_0(x) - \sum_{i=1}^{l}\gamma_i \cdot f_i(x) = \phi_0(x; \gamma = \left[\gamma_1, \cdots, \gamma_l\right]^T)\,. \qquad (4.4.12)$$

Following the same principle, the terms on the left hand side are expanded by the derivatives of $\phi_0(x;\gamma)$, which means

$$G(x) = L_\infty^{-1}(\nabla)\left(\phi_0(x;\gamma)\right) = L_\infty^{-1}\left(\nabla\right)\left(\delta(x) * \phi_0(x;\gamma)\right), \qquad (4.4.13)$$

and the convolution equation turns out as

$$\phi_0(x;\gamma) = C(x) * G(x) = L_\infty(\nabla)\big(\delta(x)\big) * L_\infty^{-1}(\nabla)\big(\phi_0(x;\gamma)\big) =$$
$$= \Big[L_\infty(\nabla)\big(\delta(x)\big) * L_\infty^{-1}(\nabla)\big(\delta(x)\big)\Big] * \phi_0(x,\gamma).$$

Notice that γ is the vector of constant coefficients; they do not have any interaction with the differential operator $L_\infty^{-1}(\nabla)$. Therefore, through the Bochner-Martin algebra that led to the equations (4.4.11), the operator $L_\infty^{-1}(\nabla)$ can still be determined as above, and the Green's function will have the representation (4.4.13) involving $\phi_0(x;\gamma)$ where γ undetermined yet.

The coefficient vector γ is now to be derived from equations (4.3.3b) which, after separating $\phi_0(x;\gamma)$ again, read:

$$\sum_{i=1}^{l}\left[\int_D f_j(x)\frac{\partial f_i(x)}{\partial x}dx + \cdots + \int_D f_j(x)\frac{\partial^n f_i(x)}{\partial x^n}dx\right]\gamma_i =$$

$$= \int_D f_j(x) L_\infty^{-1}(\nabla)\big(C_0(x)\big)dx - f_j(x_0), \qquad \text{for } j = 1,\cdots,l. \tag{4.4.14}$$

4.5 Solving the Fredholm Integral Equation in $\mathbb{R}^d$

Most often a random field $Z(x)$ will be defined in the case where $x \in D \subset \mathbb{R}^d$ and d>1; the theorem from the previous section, therefore, needs to be further generalized. The operator of equation (4.4.1), namely

$$L_n(\nabla) = \alpha_0(x) + \nabla_{\alpha_1} + \cdots + \nabla_{\alpha_{n-1}}^{n-1} + \nabla_{\alpha_n}^{n},$$

would have the terms in ∇ generalized via

$$\begin{aligned}
\nabla_{\alpha_1} &= \alpha_1(x)\frac{\partial}{\partial x_1} + \cdots + \alpha_d(x)\frac{\partial}{\partial x_d} \\
\nabla_{\alpha_2}^2 &= \alpha_{1,1}(x)\frac{\partial^2}{\partial x_1^2} + \cdots + \alpha_{d,d}(x)\frac{\partial^2}{\partial x_d^2} + \\
&\quad + \alpha_{1,2}(x)\frac{\partial^2}{\partial x_1 \partial x_2} + \cdots + \alpha_{d,d-1}(x)\frac{\partial^2}{\partial x_{d-1}\partial x_d}, \\
\nabla_{\alpha_3}^3 &= \alpha_{1,1,1}(x)\frac{\partial^3}{\partial x_1^3} + \cdots + \alpha_{1,1,2}(x)\frac{\partial^3}{\partial x_1^2 \partial x_2} + \\
&\quad + \alpha_{1,2,3}(x)\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} + \cdots + \alpha_{d-2,d-1,d}(x)\frac{\partial^3}{\partial x_{d-2}\partial x_{d-1}\partial x_d}, \quad \text{etc}
\end{aligned} \tag{4.5.1}$$

Likewise, the inverse operator from equation (4.4.4), namely

$$L_n^{-1}(\nabla) = \beta_0(x) + \nabla_{\beta_1} + \cdots + \nabla_{\beta_n}^{n},$$

will have the terms in ∇ generalized via

$$\nabla_{\beta_1} = \beta_1(x)\frac{\partial}{\partial x_1} + \cdots + \beta_d(x)\frac{\partial}{\partial x_d}$$

$$\nabla^2_{\beta_2} = \beta_{1,1}(x)\frac{\partial^2}{\partial x_1^2} + \cdots + \beta_{\mathrm{d},\mathrm{d}}(x)\frac{\partial^2}{\partial x_{\mathrm{d}}^2} + \\ + \beta_{1,2}(x)\frac{\partial^2}{\partial x_1 \partial x_2} + \cdots + \beta_{\mathrm{d}\text{-}1,\mathrm{d}}(x)\frac{\partial^2}{\partial x_{\mathrm{d}\text{-}1}\partial x_{\mathrm{d}}}$$
$$\nabla^3_{\beta_3} = \beta_{1,1,1}(x)\frac{\partial^3}{\partial x_1^3} + \cdots + \beta_{1,1,2}(x)\frac{\partial^3}{\partial x_1^2 \partial x_2} + \\ + \beta_{1,2,3}(x)\frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} + \cdots + \beta_{\mathrm{d},\mathrm{d}-1,\mathrm{d}-2}(x)\frac{\partial^3}{\partial x_{\mathrm{d}}\partial x_{\mathrm{d}-1}\partial x_{\mathrm{d}-2}} \qquad (4.5.2)$$

By extending the notations developed so far, the Kriging equation (4.3.2c) in the form of (4.4.7) will read

$$C_0(x) = C(x) * G(x) = L_\infty(\nabla)\big(\delta(x)\big) * L_\infty^{-1}(\nabla)\big(C_0(x)\big) \approx$$
$$\approx L_n(\nabla)\big(\delta(x)\big) * L_n^{-1}(\nabla)\big(C_0(x)\big) = \left(\sum_{q=0}^{n} \nabla_\alpha^q \delta(x)\right) * \left(\sum_{p=0}^{n} \nabla_\beta^p \delta(x)\right) * C_0(x),$$

using the approximate identity (4.4.8a), namely

$$\left(\sum_{q=0}^{n} \nabla_\alpha^q \delta(x)\right) * \left(\sum_{p=0}^{n} \nabla_\beta^p \delta(x)\right) \approx \delta(x).$$

Once again, the vector of coefficients $\beta \equiv \{\beta_1, \cdots, \beta_{\mathrm{d}}, \beta_{1,1}, \cdots \beta_{\mathrm{d}\text{-}1,\mathrm{d}}, \cdots\}$ is unknown; in contrast, the coefficient vectors $\alpha \equiv \{\alpha_1, \cdots \alpha_{\mathrm{d}}, \alpha_{1,1}, \cdots \alpha_{\mathrm{d},\mathrm{d}-1}\}$ has to be determined from the covariance function. Since Taylor's expansion can be applied on a d-dimensional domain, the d corresponding covariance function $C(x)$, within the convolution operator, can be expanded into

$$C(x) * G(x) = \left\{\int_D \cdots \int \delta(x - x_u) C(x_u) dx_u\right\} * G(x) =$$
$$= \left\{\int_D \cdots \int \left(\sum_{n1} \cdots \sum_{n\mathrm{d}} \frac{x_{u1}^{n1} \cdots x_{u\mathrm{d}}^{n\mathrm{d}}}{n1! \cdots n\mathrm{d}!} \delta_{x_1,\cdots,x_{\mathrm{d}}}^{\mathrm{n}1,\cdots,\mathrm{n}\mathrm{d}}(x)\right) C(x_u) dx_{u1} \cdots dx_{u\mathrm{d}}\right\} * G(x) =$$
$$= \left\{\sum_{n1} \cdots \sum_{n\mathrm{d}} \left[\delta_{x_1,\cdots,x_{\mathrm{d}}}^{\mathrm{n}1,\cdots,\mathrm{n}\mathrm{d}}(x) \int_D \cdots \int \frac{x_{u1}^{n1} \cdots x_{u\mathrm{i}\mathrm{d}}^{n\mathrm{d}}}{n1! \cdots n\mathrm{d}!} C(x_u) dx_{u1} \cdots dx_{u\mathrm{d}}\right]\right\} * G(x). \qquad (4.5.3)$$

As a result, the coefficients can be determined by the (generalized) *moment generator*

$$\alpha_{n1\cdots n\mathrm{d}} = \frac{1}{n1! \cdots n\mathrm{d}!} \int_D \cdots \int x_1^{n1} \cdots x_{\mathrm{d}}^{n\mathrm{d}} C(x) dx_1 \cdots dx_{\mathrm{d}}. \qquad (4.5.4)$$

Once again, after computing the vector α, the coefficient vector β can be determined sequentially from a similar scheme as in (4.4.11). Starting from the system

$$\alpha_0\beta_0\delta(x)+$$
$$+\left(\alpha_0\beta_{x1}+\alpha_{x1}\beta_0\right)\delta_{(x_1)}(x)+\cdots+\left(\alpha_0\beta_{xd}+\alpha_{xd}\beta_0\right)\delta_{(x_d)}(x)+$$
$$+\left(\alpha_0\beta_{x1,x1}+\alpha_{x1}\beta_{x1}+\alpha_{x1,x1}\beta_0\right)\delta_{(x_1,x_1)}(x)+\cdots+$$
$$+\left(\alpha_0\beta_{x1,x2}+\alpha_{x1}\beta_{x2}+\alpha_{x1,x2}\beta_0\right)\delta_{(x_1,x_2)}(x)+\cdots=\delta(x),$$

we would develop the following equations

$$\begin{aligned}
1&=\alpha_0\beta_0,\\
0&=\alpha_0\beta_{x_1}+\alpha_0\beta_{x1},\\
&\vdots\\
0&=\alpha_0\beta_{x1,x1}+\alpha_{x1}\beta_{x1}+\alpha_{x1,x1}\beta_0,\\
&\vdots\\
0&=\alpha_0\beta_{x1,x2}+\alpha_{x1}\beta_{x2}+\alpha_{x1,x2}\beta_0,\\
&\vdots
\end{aligned}\qquad(4.5.5)$$

An example for the case $x\in D\subset\mathbb{R}^2$ will be demonstrated in Section 6.1.

4.6 Convolution Inverse in Fourier Transform

In the previous Section 4.3, the task was indentified to solve for the Green's function in the Fredholm integral equation (4.3.2c). The problem, however, can be solved by the Fourier transform as well, that is the integral with the negative transform kernel of the harmonic functions

$$e^{-ix^T\omega}=\cos x^T\omega-i\sin x^T\omega,\qquad(4.6.1)$$

where $\omega=[\omega_1,\cdots,\omega_d]^T$ denotes the transformed vector of frequency variables in the spectral domain, and $x=[x_1,\cdots,x_d]^T$ denotes the vector of variables in the original space-time domain. The Fourier transform of a d-dimensional function $h(x)\in\mathbb{R}^{\mathrm{d}}$ is defined, in analogy to 3.3.1, by the d-dimensional integral

$$H([\omega_1,\cdots,\omega_{\mathrm{d}}]^T)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}h([x_1,\cdots,x_{\mathrm{d}}]^T)e^{-i(\omega_1x_1+\cdots+\omega_{\mathrm{d}}x_{\mathrm{d}})}dx_1\cdots dx_{\mathrm{d}}=\mathsf{F}\{h\}(\omega),\qquad(4.6.1a)$$

where the integral with exponential kernel can be operated on, layer by layer, with respect to each individual variable $x_1\to\omega_1,\cdots,x_{\mathrm{d}}\to\omega_{\mathrm{d}}$.

As for most problems involving a convolution form, by the transformation into the Fourier (spectral) domain its manipulation will be simplified. As defined in equation (4.3.2d), we denote the convolution inverse of the covariance kernel by $R(x)$; therefore, the convolution inverse can be transformed into the Fourier space as

$$\begin{aligned}
&C(x)*R(x)=\delta(x)\Rightarrow\\
&\Rightarrow\mathsf{F}\{C\}(\omega)\cdot\mathsf{F}\{R\}(\omega)=\mathsf{F}\{\delta\}(\omega)=1,
\end{aligned}\qquad(4.6.2)$$

by exploiting the Fourier transform pair $\delta(x)\leftrightarrow1$ (Arsac, 1966, p.95). Some computational advantages would be provided by this form. Observe that in equation (4.6.2), the convolution turns into a product in the spectral domain and, furthermore,

$C(x)$ as well as $R(x)$ are both mappings from the domain D with dimension d into a one-dimensional domain. Therefore, the solution of the unknown function $G(x)$ can be given the simple form

$$G(x) = R(x) * C_0(x) = \mathcal{F}^{-1}\left\{\frac{1}{\mathcal{F}\{C\}(\omega)}\right\}(x) * C_0(x)\,. \tag{4.6.3}$$

If the rational term inside the parentheses results in some linear form of polynomials in the spectral domain, then an identity deduced from the Fourier transform pair can easily transform the inverse back to the original space-time domain x. Suppose that $H_k(\omega_k; x_{\setminus k})$ (where "\" means "exclude") and $h(x_1, \cdots, x_d)$ form a Fourier transform pair in terms of the k^{th} variable, defined by

$$H_k(\omega_k; x_{\setminus k}) \equiv H_k(x_1, \cdots, x_{k-1}, \omega_k, x_{k+1}, \cdots, x_d) = \int_{-\infty}^{\infty} e^{-ix_k\omega_k} h(x) dx_k\,; \tag{4.6.4}$$

accordingly, the transform pair for the first k variables is defined by

$$\begin{aligned} &H_{1,\cdots,k}(\omega_1, \cdots, \omega_k; x_{k+1}, \cdots, x_{\mathrm{d}}) = \\ &= \int_{-\infty}^{\infty} e^{-ix_k\omega_k} H_{1,\cdots,k-1}(\omega_1, \cdots, \omega_{k-1}; x_k, \cdots, x_{\mathrm{d}}) dx_k = \\ &= \int_{-\infty}^{\infty} e^{-ix_k\omega_k} \left[\int_{-\infty}^{\infty} e^{-ix_{k-1}\omega_{k-1}} \cdot H_{1,\cdots,k-2}(\omega_1, \cdots, \omega_{k-2}; x_{k-1}, \cdots, x_{\mathrm{d}}) dx_{k-1} \right] dx_k = \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{(-x_1\omega_1 \cdots - x_k\omega_k)i} \cdot h(x_1, \cdots, x_{\mathrm{d}}) dx_1 \cdots dx_{\mathrm{k}}\,. \end{aligned} \tag{4.6.5}$$

Now for a derivative defined on the first variable, the transform in the uni-variate case is given by (Arsac, 1966, pp.39-41) :

$$\mathcal{F}_{\omega_1}\left\{\frac{\partial^{n1} h(x_1, \cdots, x_{\mathrm{d}})}{\partial x_1^{\,n1}}\right\} = (i\omega_1)^{n1} H_1(\omega_1; x_2, \cdots, x_{\mathrm{d}})\,. \tag{4.6.6}$$

By re-arranging the operators so that only related terms are left as operands, the equation (4.6.5) can be generalized to the derivative defined on a second variable by

$$\begin{aligned} &\mathcal{F}_{\omega_1,\omega_2}\left\{\frac{\partial^{n1}\partial^{n2} h(x_1, \cdots, x_{\mathrm{d}})}{\partial x_1^{n1}\partial x_2^{n2}}\right\} = \int_{-\infty}^{\infty} e^{-ix_2\omega_2} \frac{\partial^{n2}}{\partial x_2^{n2}} \left[\int_{-\infty}^{\infty} e^{-ix_1\omega_1} \frac{\partial^{n1} h(x)}{\partial x_1^{n1}} dx_1 \right] dx_2 = \\ &= (i\omega_1)^{n1} \int_{-\infty}^{\infty} e^{-ix_1\omega_1} \frac{\partial^{n2}}{\partial x_2^{n2}} [H_1(\omega_1; x_1, \cdots, x_{\mathrm{d}})] dx_2 = \\ &= (i\omega_1)^{n1} (i\omega_2)^{n2} H_{1,2}(\omega_1, \omega_2; x_2, \cdots, x_{\mathrm{d}})\,. \end{aligned} \tag{4.6.7}$$

By repeating the same manipulation, the case involving only one variable from equation (4.6.6) can be eventually generalized into the case of d variables, leading to

$$\mathcal{F}_{\omega_1,\cdots,\omega_{\mathrm{d}}}\left\{\frac{\partial^{n1} \cdots \partial^{n\mathrm{d}} h(x)}{\partial x_1^{n1} \cdots \partial x_{\mathrm{d}}^{n\mathrm{d}}}\right\} = (i)^{n1+\cdots+n\mathrm{d}} \omega_1^{n1} \cdots \omega_{\mathrm{d}}^{n\mathrm{d}} H_{1,\cdots,\mathrm{d}}(\omega)\,, \tag{4.6.8}$$

where the integers $n1,\ldots, n\mathsf{d}$ are equal or larger than zero.

By introducing the convolution with Dirac's delta function $\delta(x)$ the identity (4.6.8) is further re-shaped into

$$F_{\omega_1,\cdots,\omega_d}\left\{\frac{\partial^{n1}\cdots\partial^{nd}h(x)}{\partial x_1^{n1}\cdots\partial x_d^{nd}}\right\}=$$

$$=F_{\omega_1,\cdots,\omega_d}\left\{\frac{\partial^{n1}\cdots\partial^{nd}}{\partial x_1^{n1}\cdots\partial x_d^{nd}}[\delta(x)*h(x)]\right\}=F_{\omega_1,\cdots,\omega_d}\left\{\left[\frac{\partial^{n1}\cdots\partial^{nd}}{\partial x_1^{n1}\cdots\partial x_d^{nd}}\delta(x)\right]*h(x)\right\}=$$

$$=F\left\{\left[\frac{\partial^{n1}\cdots\partial^{nd}}{\partial x_1^{n1}\cdots\partial x_d^{nd}}\delta(x)\right]\right\}(\omega)\cdot H_{1,\cdots,d}(\omega)=(i)^{n1+\cdots+nd}\omega_1^{n1}\cdots\omega_d^{nd}H_{1,\cdots,d}(\omega), \qquad (4.6.9)$$

or

$$\left[\frac{\partial^{n1}\cdots\partial^{nd}}{\partial x^{n1}\cdots\partial x_d^{nd}}\delta(x)\right]*h(x)=F^{-1}\left\{(i)^{n1+\cdots+nd}\omega_1^{n1}\cdots\omega_d^{nd}\right\}(x)*h(x). \qquad (4.6.9a)$$

As (4.6.9a) is valid for any arbitrary function $h(x)$, we obtain, for the Dirac's delta function, the following Fourier transform pair

$$\left[\frac{\partial^{n1}\cdots\partial^{nd}}{\partial x^{n1}\cdots\partial x_d^{nd}}\delta(x)\right]\leftrightarrow(i)^{n1+\cdots+nd}\omega_1^{n1}\cdots\omega_d^{nd}. \qquad (4.6.10)$$

Example: A well known covariance model in $\mathbb{R}^1$, the exponential covariance, can serve as a good example:

$$C(x)=\beta\exp\left(-|x|/\alpha\right)$$

which has the Fourier transform pair (Oberhettinger, 1990)

$$F\{C\}(\omega)=F_C(\omega)=\frac{2\alpha\beta^2}{1+\alpha^2\omega^2}$$

In view of equation (4.3.2d), the Green's function $G(x)$ can be written in the form of

$$G(x)=R(x)*C(x)=$$

$$=F^{-1}\left\{\frac{1}{F_C(\omega)}\right\}(x)*C(x)=F^{-1}\left\{\frac{1+\alpha^2\omega^2}{2\alpha\beta^2}\right\}(x)*C(x)=$$

$$=\frac{1}{2\alpha\beta^2}\left[F^{-1}\{1\}(x)-\alpha^2F^{-1}\left\{(i\omega)^2\right\}(x)\right]*C(x)= \qquad (4.6.11)$$

$$=\frac{1}{2\alpha\beta^2}\left[\delta(x)-\alpha^2\delta''(x)\right]*C(x)=\frac{1}{2\alpha\beta^2}[C(x)-\alpha^2C''(x)]. \qquad (4.6.12)$$

The transform relation between polynomial terms in the spectral domain and derivative terms in the original domain had been provided by Arsac (1966); here it was further generalized to the d-dimensional case through equation (4.6.4)-(4.6.10).

As in the majority of cases in reality, when $1/F_C(\omega)$does not turn out to be in polynomial form and no analytical form of the inverse Fourier transform is available, one can still approximate the fractional term $1/F_C(\omega)$ into a polynomial through a power series expansion for which the inverse Fourier transform can be analytically determined term by term.

As far as a spatial-temporal covariance function is concerned, the temporal variable, usually defined in $\mathbb{R}^1$, can be regarded as an additional variable and calculated separably in the Fourier transform. For a spatial-temporal covariance function $C(h;u)$ defined on a 2-D space, its Fourier transform may be denoted by $F\{C\}(\omega;\tau)$; then the formalism of inversion in this section still applies. The convolution inverse covariance function $R(h;u)$, where $C(h;u) * R(h;u) = \delta(h;u)$, would be

$$R(h;u) = F^{-1}\left\{\frac{1}{F\{C\}(\omega;\tau)}\right\}. \tag{4.6.13}$$

4.7 Normalization by Scaling

As concluded in Chapter 2, any predictor that ought to fulfill the unbiasedness criterion would have the property

$$\mu = E\{Z_0\} = E\{\tilde{z}_0\} = E\left\{\int_D G(x)z(x)dx\right\} = \mu\int_D G(x)dx, \tag{4.7.1}$$

which leads to the constraint

$$\int_D G(x)ds = 1 \tag{4.7.2}$$

If G(x) is represented by (4.4.5), the constraint (4.7.2) implies

$\beta_0 = 1$.

On the other hand, the zero order coefficient α_0 for the covariance function in (4.4.9) must comply

$\alpha_0\beta_0 = 1$,

in observance of the identity (4.4.11). Consequently, one concludes that β_0 has to be 1 as well, hence the normalization condition of covariance function

$$\alpha_0 = \int_D C(x)dx = 1 \tag{4.7.3}$$

has to be fulfilled.

Incidentally, this conclusion can be applied to the case in $\mathbb{R}^d$ by observing the equation (4.5.5).

4.8 Positive-Definiteness of the Green's Function

Based on equation (4.6.3), the Green's function can be expressed by

$$G(x) = R(x) * C_0(x) = F^{-1}\left\{\frac{1}{F\{C\}(\omega)}\right\}(x) * C_0(x) = F^{-1}\{r(\omega)\}(x) * C_0(x) \tag{4.8.1}$$

where $r(\omega)$ denotes the Fourier transform of $R(x)$. If the positive-definite function $C(x)$ is continuous and smooth, then the following three statements can be ensured by Bochner's theorem (Bochner, 1955):

1. The function $r(\omega)$ is guaranteed to be a smooth, continuous, all positive function, because the transform $F\{C\}(\omega)$ is a positive function, and the corresponding function $r(\omega) := 1/F\{C\}(\omega)$ as well.
2. The backward transformed $F_C^{-1}\{r(\omega)\}(x)$ thus is a positive-definite function.
3. $F\{C\}(\omega)$ is a continuous, positive function because $C(x)$ is known as a positive-definite function. In the mean time, $F\{F^{-1}\{r(\omega)\}(x)\}(\omega)$ is also a positive function, because of result 2.

Consequently, it makes the spectral representation of $G(x)$, namely

$$F\{G(x)\}(\omega) = F\{F^{-1}\{r(\omega)\}(x)\}(\omega) \cdot F\{C_0(x)\}(\omega), \tag{4.8.2}$$

the product of two all positive functions. Once again by Bochner's theorem, $G(x)$ is ensured to be a positive-definite function.

Proposition 4.6.1: A Green's function $G(x)$ is a positive-definite function since both covariance functions $C(x)$ and $C_0(x)$ are positive-definite functions.

This conclusion provides a new way to create covariance functions. Since the Green's function $G(x)$ fulfills the requirement of positive-definiteness, as long as $G(x)$ also fulfills the criteria 1, 2 and 3 in Section 3.1, it will qualify as a new covariance model.

4.9 Convolution Inverse in Discrete Fourier Transform

In session 4.4, the computation of the inversion in the Spectral domain looks very promising because the analytical form of a covariance function is preserved all the way to the inverse solution, in terms of its parameters. However, such doctrine does require the form of the inverse Fourier transform pair to be analytical, which will be available only for very few special examples rather than for the general case. To overcome this technical problem, numerical solutions may be obtained from a discrete Fourier transform, such as DFT(Bachman et al., 2000, chapter 6), FFT(Loan, 1992) and are supported by efficient computer programs. The inverse transform in equation (4.8.1), with a properly selected interval $[a_1, b_1]\times\ldots\times[a_d, b_d]$ will be approximated by

$$F^{-1}\left\{\frac{1}{F\{C\}(\omega)}\right\}(x) = F^{-1}\{r(\omega)\}(x)$$
$$\approx \sum_{k1=0}^{n1-1}\cdots\sum_{kd=0}^{kd-1} r([\omega_{k1},\cdots\omega_{kd}]^T)e^{-2\pi k1/n1}\cdots e^{-2\pi kd/nd}\Delta x_1\cdots\Delta x_d .$$

For many covariance models, the spectral representation in analytical form is indeed available because the positive-definiteness has to be established this way. Therefore, the inverse transform back to the original domain is the concern, where the discrete Fourier transform can serve well. In other cases, the analytical form of a covariance function is only available in the spectral representation. This will not cause much problems since manipulations are primarily carried out in the spectral domain and the final transform back to the original domain can be approximated using discrete techniques.

4.10 Green's Function and Parameter Estimation of the Random Field

In this section, we are going to have an outlook on the approach to fit random field data into its parametric model. By way of the approach discussed early in this chapter (4.4, 4.5, and 4.6), Green's function can be represented by the same parameters with the covariance function. Therefore, when combined with conditionally specified parametric model of the random field, those parameters is able to be estimated from the observed data directly.

The Section 2.3 concludes that the joint distribution of a random field can be factorized into terms of suitable conditional distributions. In the case of a Gaussian density, the random field Z has the joint density

$$\underset{n\times 1}{Z} \sim N\left(\mu,(I-G)^{-1}D\right),$$

in accordance with (2.3.4), $D \equiv Diag\left(\sigma_1^2,\cdots,\sigma_n^2\right)$ and the $n\times n$ matrix $G \equiv \left(g_{ij}\right)$ is unknown parameters of this random field, which can be estimated by the maximization of the negative loglikelihood when the observed data $\underset{n\times 1}{z}$ is provided. In Cressie (1991, p467), the negative loglikelihood of the model for a site x_0 is given by

$$L(\mu,\sigma_0^2,G=\{g_{0j}\}) = \left(n/2\right)\log(2\pi\sigma_0^2) - \left(1/2\right)\log(|I-G|) + \\ + \left(1/2\right)\underset{n\times 1}{(Z-\mu)}^T(I-G)\underset{n\times 1}{(Z-u)}/\sigma_0^2 . \qquad (4.10.1)$$

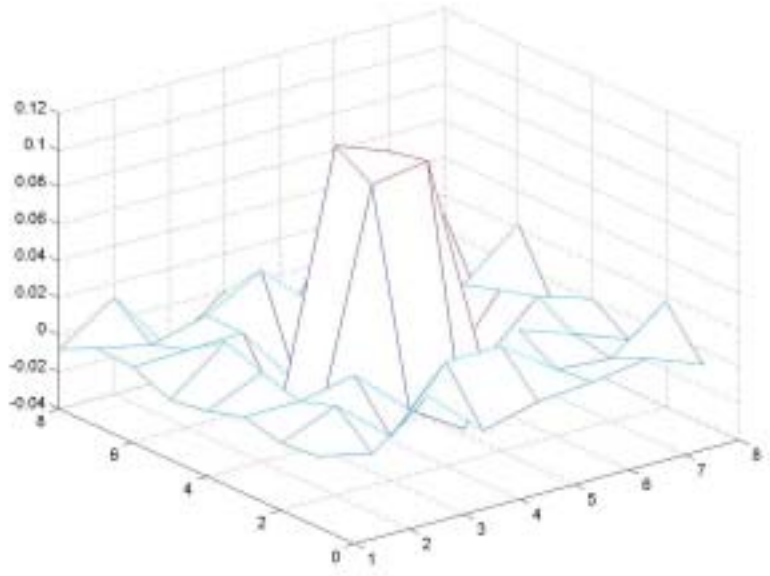


Figure 4.6: The estimated coefficients of $\hat{G} \equiv \left(\hat{g}_{ij}\right)$

To determine the unknown $G \equiv \left(g_{0j}\right)$ through (4.10.1), additional information can often be fed through the parameters μ, σ_0^2, and G on the respective applications. For example, the mean is often either provided, or at least assumed to be constant; constraints are imposed onto the matrix $G \equiv \left(g_{0j}\right)$, based on assumption about the field such as isotropy. Afterwards, the remaining unknowns will be estimated by minimizing the negative loglikehood. Figure 4.6 shows a computation of example of the negative loglikehood estimator, where the estimated parameters in $\underset{8\times 8}{\hat{G}} \equiv \left(\hat{g}_{0j}\right)$, constrained by isotropy condition, are plotted.

Meanwhile, in accordance with (2.3.2), the conditional mean at x_0 has a linear form involving those undetermined coefficients g_{oj} :

$$E\left\langle z(x_0)\middle| z(x_j): j \neq 0\right\rangle = \mu_0 + \sum_{j=1}^{n} g_{0j}\left(z\left(x_j\right) - \mu_j\right) = \mu_0 + \sum_{j=1}^{q} g_{0j} \cdot \left(z\left(x_j\right) - \mu_j\right).$$

Such a linear form shows analogy with linear predictor discussed in Chapter 2 and 4; By the comparison with (4.1.4), the coefficients g_{ij} is associated with the Kriging coefficients, or discretized Green's function if the condition mean is considered as the inhomogeneously Best Linear predictor. Moreover, the conditional variance is given by

$$E\left\{z_0 - E\left\langle z(x_0)\middle| z(x_j): j \neq 0\right\rangle\right\} = \sigma_{x_0}^2 - \int_D C(x)G(x)dx \cong \sigma_{x_0}^2 - \underset{n\times 1}{C_{0z}^T}\, g_{0j},$$

which is the Kriging variance in (2.4.4). Such an association between the optimal predictor and the conditional density has been revealed in Cressie (1991, p.108, p.469).

Now that the unknown $G \equiv \left(g_{0j}\right)$ in (4.10.1) is considered as discretely evaluated Green's function $G(x_a)$, where x_a runs through the set of all the observation sites, we can join together the estimated matrix $\hat{G} \equiv \left(\hat{g}_{ij}\right)$ along with its dispersion matrix $D\left\{\hat{G}\right\}$, with the analytically represented Green's function $G(x_a)$. In order to simplify the notation, we can sequentially record the matrix $\hat{G} \equiv \left(\hat{g}_{ij}\right)$ over one dimension after another so the matrix $\hat{G}$ is collected by a vector in the same order as x_a, which is denoted by $vec\left\{\hat{G}\right\}$. If we further considering the $vec\left\{\hat{G}\right\}$ as observation vector, the parameters of associated function, embedded in the Green's function expression (4.4.5), can be estimated through the following Gauss-Markov model as

$$\begin{aligned} vec\left\{\hat{G}\right\} &= vecL_{\infty}^{-1}C(x = x_a;\xi) + e \\ &\approx A\xi + e \end{aligned} \tag{4.10.2}$$

$$e \sim \left(0, D\left\{vec(\hat{G})\right\}\right), \tag{4.10.3}$$

where ξ denotes the parameters of the covariance function, and A is the Jacobian matrix of the function in (4.10.2) with respect of ξ.

CHAPTER 5

EXAMPLES AND EXPERIMENTS

5.1 Green's Function for a Separable Covariance Function

The first *example* shows the representation of Green's function associated with a stationary 2-dimensional random field with a known mean. Let us suppose that the isotropic covariance function is given by

$$C(x,y;a)=\frac{1}{a^2\pi}\exp\left(\frac{-x^2-y^2}{a^2}\right). \qquad (5.1.1)$$

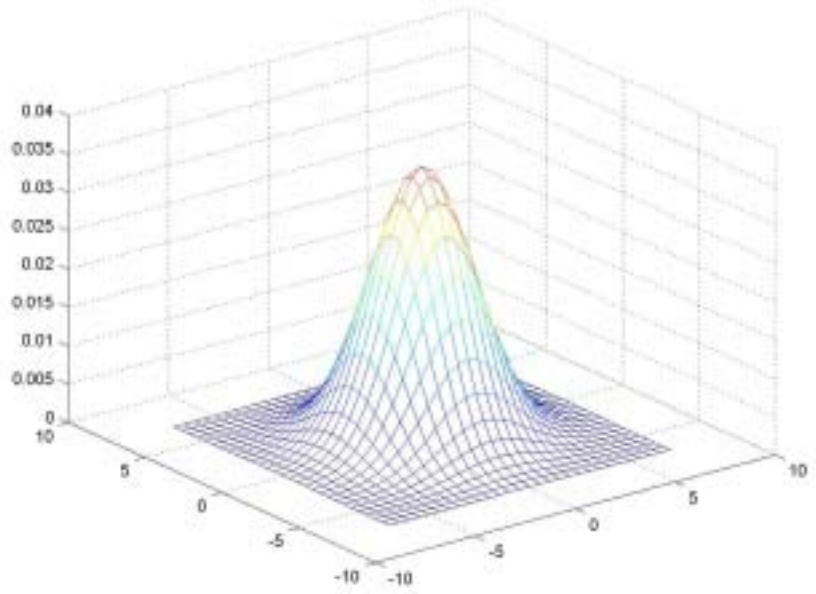


Figure 5.1: The covariance function of equation (5.1.1)

This 2-D covariance function is expanded as in (4.5.3), yielding

$$C(x,y)=\iint\delta(x-x_u,y-y_u)C(x_u,y_u)dx_u dy_u =$$

$$=\iint\left[\delta(x,y-y_u)-\frac{\partial\delta(x,y-y_u)}{\partial x}x_u+\frac{\partial^2\delta(x,y-y_u)}{2!\partial x^2}x_u^2-\cdots\right]C(x_u,y_u)dx_u dy_u$$

$$=\iint C(x_u,y_u)\left\{\left[\delta(x,y)-\frac{\partial\delta(x,y)}{\partial y}y_u+\frac{\partial^2\delta(x,y)}{2!\partial y^2}y_u^2-\frac{\partial^3\delta(x,y)}{3!\partial y^3}y_u^3+\cdots\right]-\right.$$

$$-\left[\frac{\partial\delta(x,y)}{\partial x}-\frac{\partial^2\delta(x,y)}{\partial x\partial y}y_u+\frac{\partial^3\delta(x,y)}{2!\partial x\partial y^2}y_u^2-\frac{\partial^4\delta(x,y)}{3!\partial x\partial y^3}y_u^3+\cdots\right]x_u+$$
$$+\frac{1}{2!}\left[\frac{\partial^2\delta(x,y)}{\partial x^2}-\frac{\partial^3\delta(x,y)}{\partial x^2\partial y}y_u+\frac{\partial^4\delta(x,y)}{2!\partial x^2\partial y^2}y_u^2-\frac{\partial^5\delta(x,y)}{3!\partial x^2\partial y^3}y_u^3+\cdots\right]x_u^2-$$
$$-\frac{1}{3!}\left[\frac{\partial^3\delta(x,y)}{\partial x^3}-\frac{\partial^4\delta(x,y)}{\partial x^3\partial y}y_u+\frac{\partial^5\delta(x,y)}{2!\partial x^3\partial y^2}y_u^2-\frac{\partial^6\delta(x,y)}{3!\partial x^3\partial y^3}y_u^3+\cdots\right]x_u^3+\cdots\Big\}dx_u dy_u . \tag{5.1.2}$$

Consequently, an operator can be written as

$$C(x,y)*=[\alpha_{00}\delta(x,y)-\alpha_{01}\delta^{(0)(1)}(x,y)-\alpha_{10}\delta^{(1)(0)}(x,y)+\alpha_{11}\delta^{(1)(1)}(x,y)+\cdots]* \tag{5.1.3}$$

where the coefficients α_{ij} are calculated from the correspondent moment generators, written in matrix form:

$\iint C(x,y)dxdy=1$	$-\iint yC(x,y)dxdy=0$	$\frac{1}{2}\iint y^2C(x,y)dxdy=\frac{a^2}{4}$	$\frac{-1}{6}\iint y^3C(x,y)dxdy=0$
$-\iint xC(x,y)dxdy=0$	$\iint xyC(x,y)dxdy=0$	$\frac{-1}{2}\iint xy^2C(x,y)dxdy=0$	$\frac{1}{3}\iint xy^3C(x,y)dxdy=0$
$\frac{1}{2}\iint x^2C(x,y)dxdy=\frac{a^2}{4}$	$\frac{-1}{2}\iint x^2yC(x,y)dxdy=0$	$\frac{1}{4}\iint x^2y^2C(x,y)dxdy=\frac{a^4}{16}$	$\frac{-1}{12}\iint x^2y^3C(x,y)dxdy=0$
$\frac{-1}{6}\iint x^3C(x,y)dxdy=0$	$\frac{1}{3}\iint x^3yC(x,y)dxdy=0$	$\frac{-1}{12}\iint x^3y^2C(x,y)dxdy=0$	$\frac{1}{36}\iint x^3y^3C(x,y)dxdy=0$

Meanwhile, the convolution inverse $R(x,y)$ of the covariance function is also written in terms of delta function derivatives, with unknown coefficient β_{ij}

For a 2-dimensional covariance function, the analog of equation (4.4.8) in conjunction with (4.5.1-2), leads first to:

$$\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\alpha_{ij}\nabla^{ij}\delta(x,y)\right)*\left(\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\beta_{pq}\nabla^{pq}\delta(x,y)\right)=\delta(x,y),$$

where ∇^{ij} denotes the partial derivative $\partial^{(i+j)}/\partial x^i\partial y^j$, and consequently to the following identities as in (4.5.5):

$\alpha_{00}\beta_{00}=1 \quad \Rightarrow \beta_{00}=1$

$\alpha_{10}\beta_{00}+\alpha_{00}\beta_{10}=0 \quad \Rightarrow \beta_{10}=0$

$\alpha_{01}\beta_{00}+\alpha_{00}\beta_{01}=0 \quad \Rightarrow \beta_{01}=0$

$\alpha_{00}\beta_{11}+\alpha_{10}\beta_{01}+\alpha_{01}\beta_{10}+\alpha_{11}\beta_{00}=0 \quad \Rightarrow \beta_{11}=0$

$\alpha_{00}\beta_{20}+\alpha_{10}\beta_{10}+\alpha_{20}\beta_{00}=0 \Rightarrow \beta_{20}=-\frac{1}{4}a^2$

$$\alpha_{00}\beta_{02}+\alpha_{01}\beta_{01}+\alpha_{02}\beta_{00}=0 \Rightarrow \beta_{02}=-\frac{1}{4}a^2$$

$$\alpha_{00}\beta_{21}+\alpha_{10}\beta_{11}+\alpha_{20}\beta_{01}+\alpha_{21}\beta_{00}+\alpha_{01}\beta_{20}=0 \qquad \Rightarrow \beta_{21}=0$$

$$\alpha_{00}\beta_{12}+\alpha_{01}\beta_{11}+\alpha_{02}\beta_{10}+\alpha_{12}\beta_{00}+\alpha_{10}\beta_{02}=0 \qquad \Rightarrow \beta_{12}=0$$

$$\alpha_{00}\beta_{22}+\alpha_{01}\beta_{21}+\alpha_{10}\beta_{12}+\alpha_{11}\beta_{11}+\alpha_{20}\beta_{02}+\alpha_{02}\beta_{20}+\alpha_{21}\beta_{01}+\alpha_{12}\beta_{10}+\alpha_{22}\beta_{00}=0$$

$$\Rightarrow \beta_{22}=\frac{1}{16}a^4$$

After the determination of coefficients β_{ij}, the Green's function has the form of

$$G(x,y;a)=\left(\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\beta_{pq}\nabla^{pq}\delta(x,y)\right)*C_0(x,y;a)=\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\beta_{pq}\nabla^{pq}C_0(s,y;a)=$$

$$=\beta_{0,0}C(x,y;a)+\beta_{2,0}\frac{\partial^2 C(x,y;a)}{\partial x^2}+\beta_{0,2}\frac{\partial^2 C(x,y;a)}{\partial y^2}+\beta_{2,2}\frac{\partial^2\partial^2 C(x,y;a)}{\partial x^2\partial y^2}+\cdots \qquad (5.1.4)$$

where

$$\frac{\partial^2 C_o(x,y;a)}{\partial x^2}=\left(\frac{-2}{a^4\pi}+\frac{4x^2}{a^6\pi}\right)\exp\left(\frac{-x^2-y^2}{a^2}\right)$$

$$\frac{\partial^2 C_o(x,y;a)}{\partial y^2}=\left(\frac{-2}{a^4\pi}+\frac{4y^2}{a^6\pi}\right)\exp\left(\frac{-x^2-y^2}{a^2}\right)$$

$$\frac{\partial^2\partial^2 C_o(x,y;a)}{\partial x^2 y^2}=\left(\frac{4}{a^6\pi}-\frac{8x^2}{a^8\pi}-\frac{8y^2}{a^8\pi}+\frac{16x^2y^2}{a^{10}\pi}\right)\exp\left(\frac{-x^2-y^2}{a^2}\right)$$

$$\beta_{0,0}=1, \qquad \beta_{2,0}=\beta_{0,2}=\frac{-a^2}{4}, \quad \beta_{2,2}=\frac{a^4}{16}$$

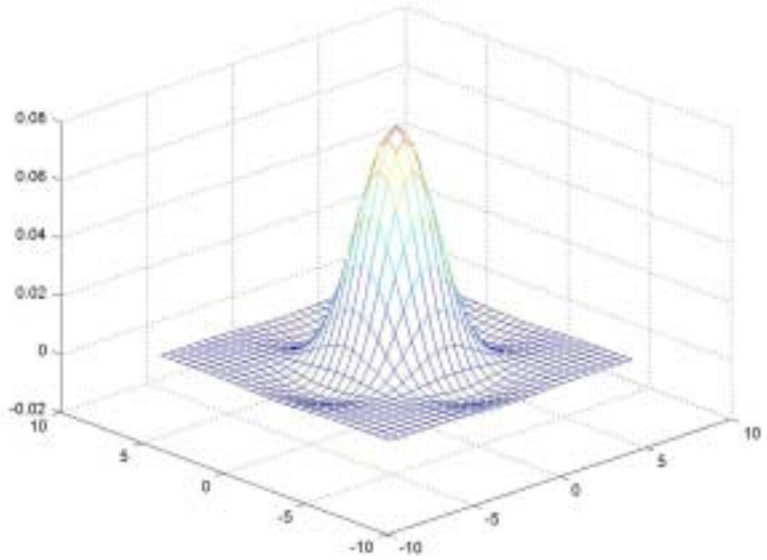


Figure 5.2: The Green's function according to equation (5.1.4)

So we have a representation of Green's function which retains the parameter of the original covariance function in the 3-D plot in figure 5.2.
The *next example* is an anisotropic covariance function, defined by

$$C(x,y)=\frac{1}{ab\pi}\exp\left(-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right) \tag{5.1.5}$$

and plotted in Figure 5.3

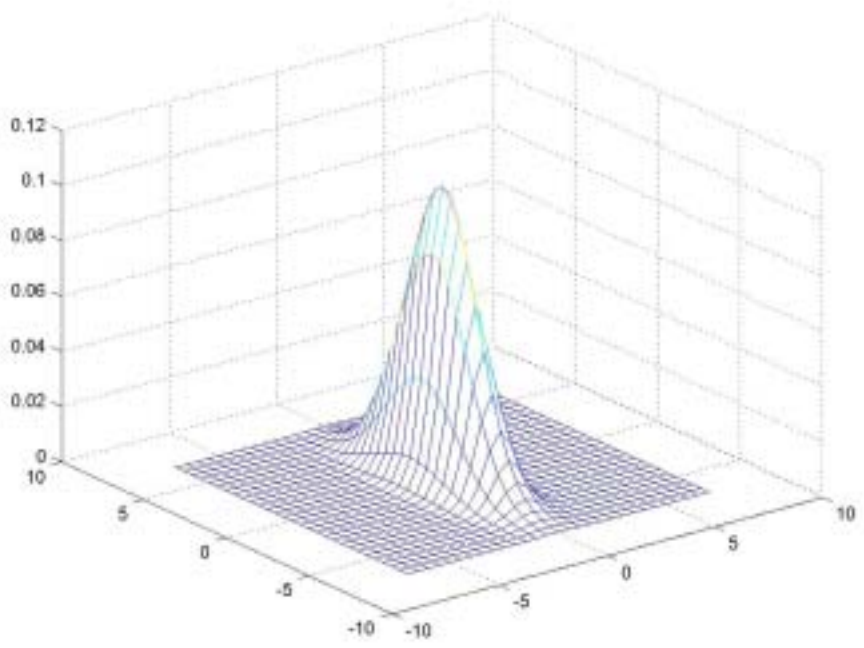


Figure 5.3: The covariance function of equation (5.1.5)

According to the equation (5.1.2), the coefficients α_{ij} are calculated from the moment generators in the following manner, listed in matrix form:

$\iint C(x,y)dxdy=1$	$-\iint yC(x,y)dxdy=0$	$\frac{1}{2}\iint y^2C(x,y)dxdy=\frac{a^2}{4}$	$\frac{-1}{6}\iint y^3C(x,y)dxdy=0$
$-\iint xC(x,y)dxdy=0$	$\iint xyC(x,y)dxdy=0$	$\frac{-1}{2}\iint xy^2C(x,y)dxdy=0$	$\frac{1}{6}\iint xy^3C(x,y)dxdy=0$
$\frac{1}{2}\iint x^2C(x,y)dxdy=\frac{b^2}{4}$	$\frac{-1}{2}\iint x^2yC(x,y)dxdy=0$	$\frac{1}{4}\iint x^2y^2C(x,y)dxdy=\frac{(ab)^2}{16}$	$\frac{-1}{12}\iint x^2y^3C(x,y)dxdy=0$
$\frac{-1}{6}\iint x^3C(x,y)dxdy=0$	$\frac{1}{6}\iint x^3yC(x,y)dxdy=0$	$\frac{-1}{12}\iint x^3y^2C(x,y)dxdy=0$	$\frac{1}{36}\iint x^3y^3C(x,y)dxdy=0$

Once again, through the expansion of (5.1.3), the coefficients β_{ij} can be calculated for $0 \le i, j \le 3$ as follows:

$$\beta_{00} = 1$$

$$\beta_{20} = -\frac{1}{4}b^2 \quad \beta_{02} = -\frac{1}{2}a^2\,\frac{1}{2!}$$

$$\beta_{2,2} = \frac{a^2 b^2}{16}$$

with the following partial derivatives of $C_0(x, y; a, b)$:

$$\frac{\partial^2 C_0(x,y;a,b)}{\partial x^2} = \frac{1}{ab\pi}\left(\frac{-2}{a^2} + \frac{4x^2}{a^4}\right)\exp\left(-\frac{x^2}{a^2} - \frac{y^2}{b^2}\right)$$

$$\frac{\partial^2 C_0(x,y;a,b)}{\partial x^2} = \frac{1}{ab\pi}\left(\frac{-2}{a^2} + \frac{4x^2}{a^4}\right)\exp\left(-\frac{x^2}{a^2} - \frac{y^2}{b^2}\right)$$

$$\frac{\partial^2 \partial^2 C_0(x,y;a,b)}{\partial x^2 y^2} = \frac{1}{ab\pi}\left(\frac{-2}{a^2} + \frac{4x^2}{a^4}\right)\left(\frac{-2}{b^2} + \frac{4y^2}{b^4}\right)\exp\left(-\frac{x^2}{a^2} - \frac{y^2}{b^2}\right)$$

The Green's function now has the form

$$G(x,y;a,b) = \left(\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\beta_{pq}\nabla^{pq}\delta(x,y)\right) * C_0(x,y;a,b) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\beta_{pq}\nabla^{pq}C_0(x,y;a,b) =$$

$$= \beta_{0,0}C_o(x,y;a,b) + \beta_{2,0}\frac{\partial^2 C_o(x,y;a,b)}{\partial x^2} +$$

$$+ \beta_{0,2}\frac{\partial^2 C_o(x,y;a,b)}{\partial y^2} + \beta_{2,2}\frac{\partial^2 \partial^2 C_o(x,y;a,b)}{\partial x^2 \partial y^2} + \cdots \qquad (5.1.6)$$

and is plotted in Figure 5.4.

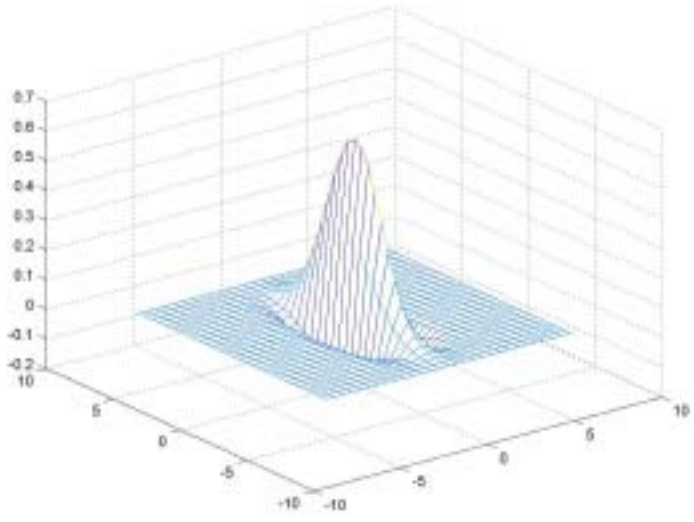


Figure 5.4: The Green's function according to (5.1.6)

5.2 Green's Function for Cressie and Huang's Covariance Function

Cressie and Huang (1999) had proposed several spatio-temporal covariance functions. The following expression is a selection from that family of functions:

$$C(h;u \mid \theta) = \sigma^2 \exp\left\{ - a^2 u^2 - b^2 \|h\|^2 - cu^2 \|h\|^2 \right\}. \qquad (5.2.1).$$

For the sake of visual illustration, the spatial variable h is assumed to be defined in a one-dimensional space. Equation (5.2.1) is a two-dimensional covariance function parameterized by the vector $\theta = \left[a, b, c, \sigma^2\right]^T$. In this case, not all the moments have an analytical integral readily available although we do not face integrability problems per se. So, approaches of numerical analysis can indeed provide a powerful tool to approximate the integrals for every moment.

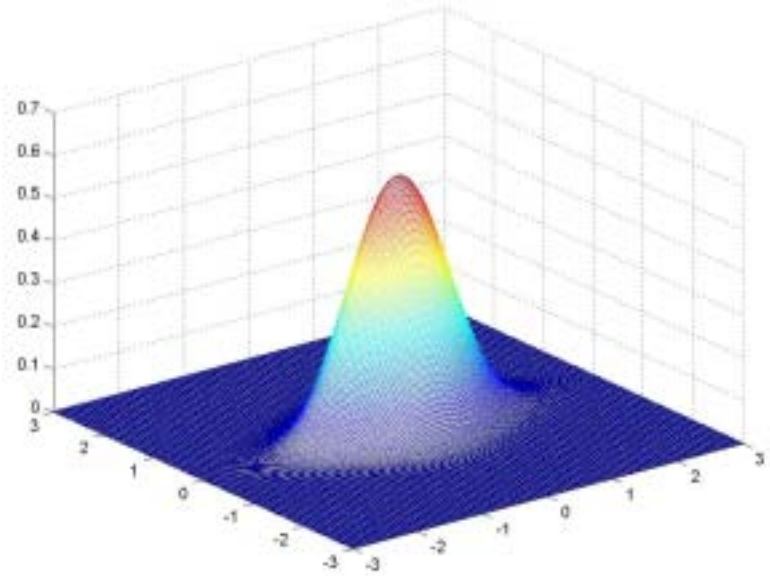


Figure 5.5: Spatial-temporal covariance function of equation (5.2.1)

Let us assume that the parameters of the covariance function are given as a^2=1.0, b^2=3.0, c=1.0, and σ^2=1.0 (Figure 5.5). Similarly to the computations in Section 5.1, the moments of the covariance foundation from equation (5.2.1) now are computed by numerical integration. Let α_{ij} denote the moment integral with power of h equal to i and power of y equal to j, for example, $\alpha_{12} = \frac{1}{2!}\iint xy^2 C dxdy$, so the moment integrals are tabulated as

$\alpha_{00} = 1$	$\alpha_{01} = 0$	$\alpha_{02} = 0.2229$	$\alpha_{03} = 0$
$\alpha_{10} = 0$	$\alpha_{11} = 0$	$\alpha_{12} = 0$	$\alpha_{13} = 0$
$\alpha_{20} = 0.0743$	$\alpha_{21} = 0$	$\alpha_{22} = 0.0135$	$\alpha_{23} = 0$
$\alpha_{30} = 0$	$\alpha_{31} = 0$	$\alpha_{32} = 0$	$\alpha_{33} = 0$

Figure 5.6 and Figure 5.7 depict the moment integrands of α_{10} and α_{22}. By the computation of moment integrals, the covariance function is transformed into a Taylor's type representation in the space spanned by its derivatives, following the same approach as in previous section. Let us truncate the terms after the third order; then the coefficients $\alpha_{i,j}$, $0 \le i, j \le 3$, will be zero, except for $\alpha_{00} = 1$, $\alpha_{20} = 0.0743$, $\alpha_{02} = 0.2229$, and $\alpha_{22} = 0.0135$

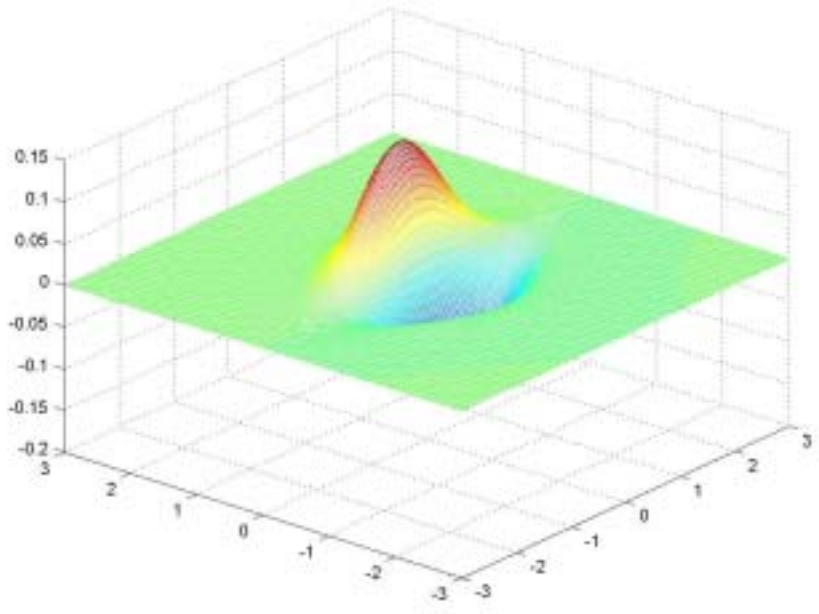


Figure 5.6: The moment function $hC(h,u,\theta)$

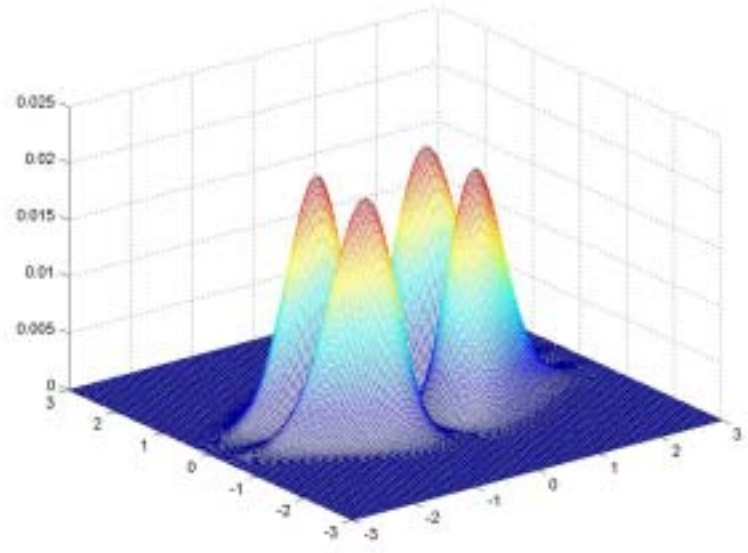


Figure 5.7: The moment function $h^2u^2C(h,u;\theta)$

Consequently, the representation of the convolution inverse $R(x,y;\theta)$is given by the coefficients $\beta_{00}=1$, $\beta_{20}=-0.0743$, $\beta_{02}=-0.2229$, and $\beta_{22}=0.0417$. The associated Green's function is then represented by combination of the partial derivatives of covariance function $C_0(h,u;\theta)$, which leads to

$$G(h;u;\theta)$$
$$=\beta_{0,0}C_0(h,u;\theta)+\beta_{2,0}\frac{\partial^2 C_0(h,u;\theta)}{\partial h^2}+\beta_{0,2}\frac{\partial^2 C_0(h,u;\theta)}{\partial u^2}+\beta_{2,2}\frac{\partial^2\partial^2 C_0(h,u;\theta)}{\partial h^2\partial u^2}+\cdots$$

letting $D_h=-2a^2-2ch^2$, $D_u=-2b^2-2cu^2$, and $\exp\left(-a^2u^2-b^2h^2-ch^2u^2\right)=e^{-P}$, the resulting derivatives are

$$\frac{\partial^2 C_0(h,u;\theta)}{\partial u^2}=\sigma^2 D_h\mathrm{e}^{-\mathrm{P}}+\sigma^2 D_h^2u^2\mathrm{e}^{-\mathrm{P}}$$
$$\frac{\partial^2 C(h,u;\theta)}{\partial h^2}=\sigma^2 D_u\mathrm{e}^{-\mathrm{P}}+\sigma^2 D_u^2h^2\mathrm{e}^{-\mathrm{P}}$$
$$\frac{1}{\sigma^2}\frac{\partial^4 C(h,u;\theta)}{\partial h^2\partial u^2}=\{-4c-8ch^2D_u+D_hD_u+D_hD_u^2h^2+32c^2h^2u^2+$$
$$-16chuD_hD_uhu-8cu^2D_h+u^2D_h^2D_u+h^2u^2D_h^2D_u^2\}e^{-P}.$$

The resulting Green's function is illustrated in figure 5.8.

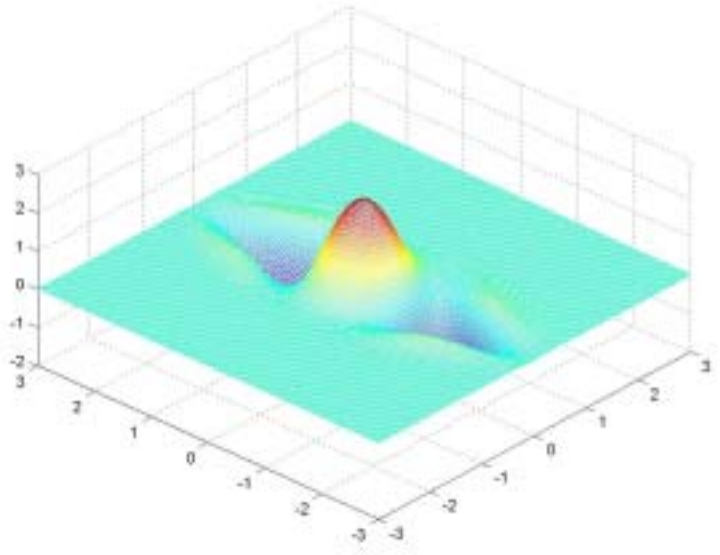


Figure 5.8: The Green's function as derived from the covariance function (5.2.1)

5.3 Green's Function for Gneiting's Covariance Function Family

Another covariance function which had been proposed by Cressie and Huang (1999) and can serve as an example of Gneiting's general form, is formulated by (5.3.1)

$$C(h;u;\theta) = \frac{\sigma^2}{\sqrt{(a^2u^2+1)}} \exp\left\{\frac{-b^2h^2}{a^2u^2+1}\right\}, \qquad \theta := \left[a, b, \sigma^2\right]^T, \tag{5.3.1}$$

and plotted in Figure 5.9.

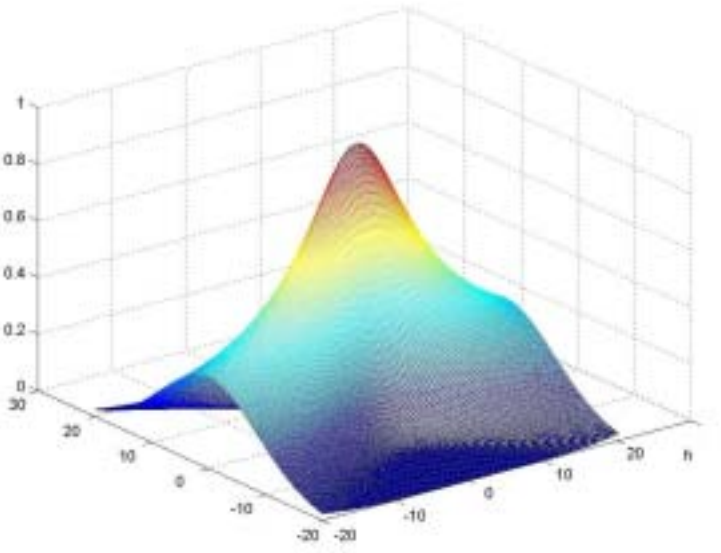


Figure 5.9: Covariance function of equation (5.3.1)

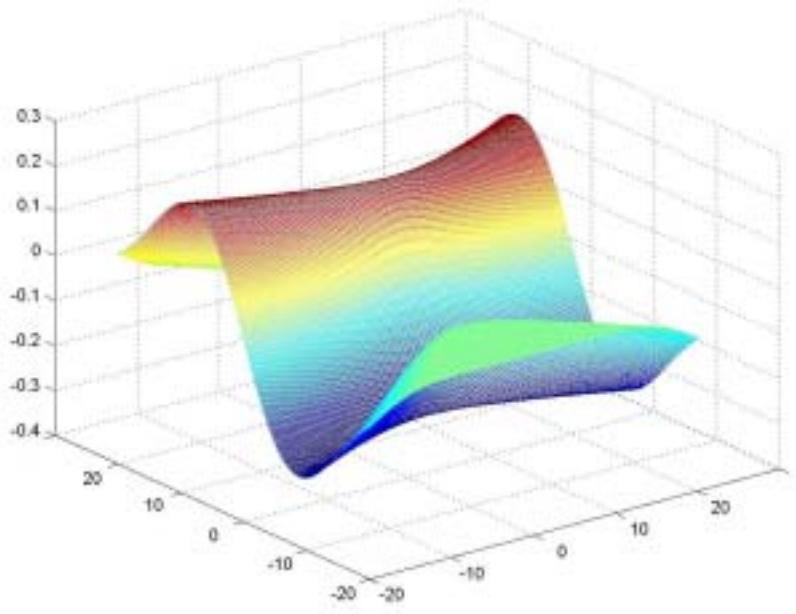


Figure 5.10: The moment function $hC(h,u;\theta)$ which leads to divergent integrals

This is a very good example where some moments do not exist. For example, the integral of the moment generator

$$\iint h \cdot C(h;u;\theta)dhdu\,,$$

does not converge for any choice of θ. Figure 5.10 illustrates this moment function, suggesting a tendency for divergence. Such a property shatters the hope to express this covariance function in the space spanned by its derivatives simply by computing the moment generators.

Alternatively, the Fredholm integral equation can be solved by the Fourier representation since the Fourier transform of this function still exists. But first, the Fourier representation of the covariance function in (5.3.1)needs to be found analytically. Cressie and Huang (1999) provided a partially complete transform with respect to h, thus it is only necessary to further integrate the transform with respect to the remaining time variable u. The Fourier representation of $C(h,u;\theta)$ from (5.3.1) is derived through

$$g(\omega;\tau) = \int \exp(iu\tau)\exp(-\omega^2 u^2)\exp(-4\varsigma u^2)\exp(-k\omega^2)du =$$

$$= \frac{2\sqrt{\pi}}{\sqrt{\omega^2+\varsigma}}\exp\{-k\omega^2\}\exp\left\{\frac{-4\pi^2\tau^2}{\omega^2+\varsigma}\right\}.$$

Which, after pointwise inversion, yields:

$$v(\omega;\tau) = \left(\frac{1}{g(\omega;\tau)}\right) = \frac{\sqrt{\omega^2+\varsigma}}{2\sqrt{\pi}}\exp(k\omega^2)\exp\left(\frac{4\pi^2\tau^2}{\omega^2+\varsigma}\right) =$$

$$= v(\omega_0 = 0;\tau_0 = 0) + \frac{\partial v(\omega_0;\tau_0)}{\partial \omega}\omega + \frac{\partial v(\omega_0;\tau_0)}{\partial \tau}\tau + \frac{\partial^2 v(\omega_0;\tau_0)}{\partial \omega \partial \tau}\omega\tau +$$
$$+ \frac{\partial^2 v(\omega_0;\tau_0)}{2!\partial \omega^2}\omega^2 + \frac{\partial^2 v(\omega_0;\tau_0)}{2!\partial \tau^2}\tau^2 + \cdots$$
$$= \frac{\sqrt{\varsigma}}{2\sqrt{\pi}} + \left(\frac{1}{2\sqrt{\pi\varsigma}} + k\sqrt{\frac{\varsigma}{\pi}}\right)\omega^2 + \left(\frac{4\pi^2}{\sqrt{\pi\varsigma}}\right)\tau^2 + \cdots$$

We notice that only terms of even power remain in the equation, which ensures that the inverse Fourier transform will be a real-valued function, formally represented as

$$F^{-1}\{v(\omega;\tau)\} = \frac{\sqrt{\varsigma}}{2\sqrt{\pi}}\delta(h;u) - \left(\frac{1}{2\sqrt{\pi\varsigma}} + k\sqrt{\frac{\varsigma}{\pi}}\right)\frac{\partial^2\delta(h;u)}{\partial h^2} + \left(\frac{4\pi^2}{\sqrt{\pi\varsigma}}\right)\frac{\partial^2\delta(h;u)}{\partial u^2} + \cdots;$$

following (4.6.3), the Green's function associated with (5.3.1) is now readily obtained as

$$G(h,u;\theta) = F^{-1}\{v(\omega;\tau)\} * C(h,u;\theta) =$$
$$= \frac{\sqrt{\varsigma}}{2\sqrt{\pi}}C(\omega;\tau) - \left(\frac{1}{2\sqrt{\pi\varsigma}} + k\sqrt{\frac{\varsigma}{\pi}}\right)\frac{\partial^2 C(h,u;\theta)}{\partial h^2} + \left(\frac{4\pi^2}{\sqrt{\pi\varsigma}}\right)\frac{\partial^2 C(h,u;\theta)}{\partial u^2} + \cdots \quad (5.3.2)$$

Illustration of the Green's function is provided in Figure 5.11; notice that the example has not been normalized to follow the scaling conditions from Section 4.7.

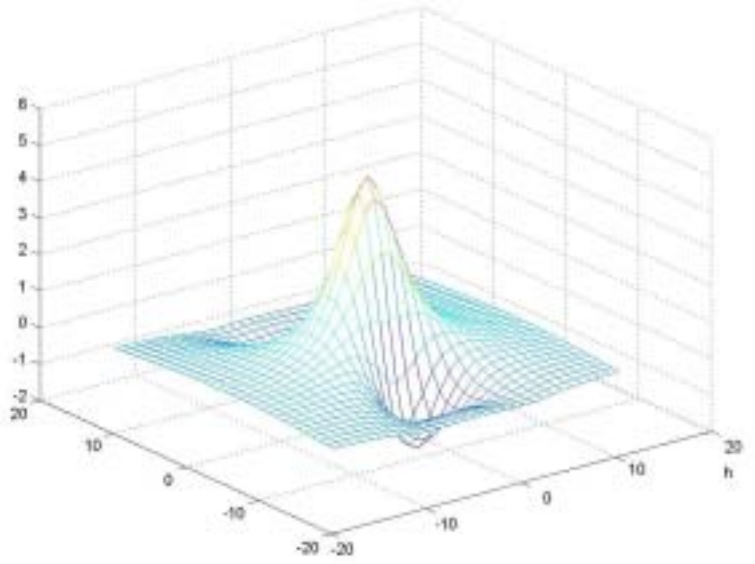


Figure 5.11: Green's function from equation (5.3.2)

5.4 Green's Function for Stein's Covariance Function Family

In his report, Stein (2003) mentioned a spatial-temporal covariance family which is represented in the spectral domain. The backward transform onto the original space-time domain for such a covariance family turns out to be quite a demanding job and only feasible with certain constraints. With the help of the convolution inverse theorem and

some numerical analysis tools, the associated Green's functions represented in the space-time domain can be computed.

One example, selected from Stein's covariance family, is formulated as

$$f_C(\omega;\tau) = \left\{c_1(a_1^2+\omega^2)^3 + c_2(a_2^2+\tau^2)\right\}^{-1}, \tag{5.4.1}$$

and plotted in Figure 5.12 as a function of the spatial variable ω, and the temporal variable τ in the spectral domain. It depends on the four parameters a_1, a_2, c_1 and c_2.

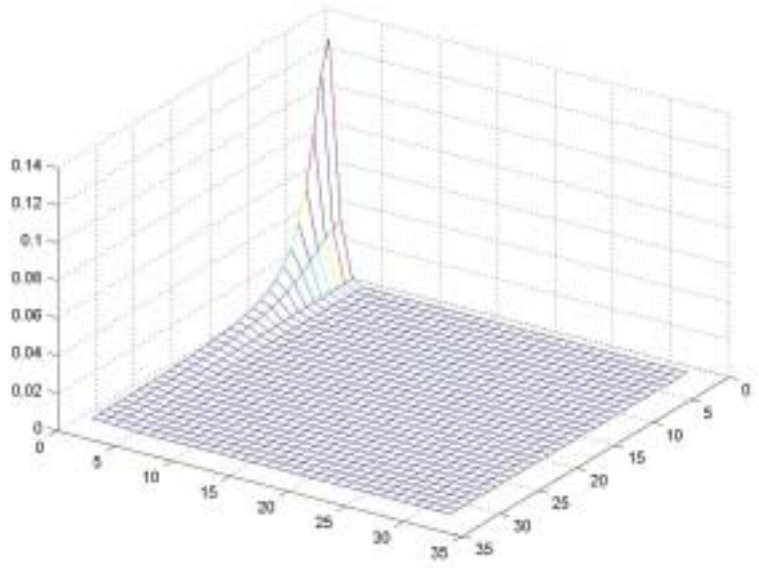


Figure 5.12 Stein's covariance function (1st quadrant)

Following the notation used before, the spatial variable *h*, and the temporal variable *u* denote the correspondent counterpart of ω and τ in original domain; that is, $F\{C(h;u)\}(\omega;\tau) = f(\omega;\tau)$. Based on the relation (4.6.3) involving the convolution inverse, the Green's function associated with (5.4.1) can be written as

$$G(h;u) = F^{-1}\left\{\frac{1}{f_C(\omega,\tau)}\right\} * C_0(h;u) \tag{5.4.2}$$

Accordingly, the convolution inverse would be formally obtained as

$$F^{-1}\left\{\frac{1}{f_C(\omega;\tau)}\right\} = F^{-1}\left\{c_1\left(a_1^6 + 3a_1^4\omega^2 + 3a_1^2\omega^4 + \omega^6\right) + c_2\left(a_2^2+\tau^2\right)\right\} =$$

$$= F^{-1}\left\{c_1\left(a_1^6 - 3a_1^4(i\omega)^2 + 3a_1^2(i\omega)^4 - (i\omega)^6\right) + c_2\left(a_2^2 - (i\tau)^2\right)\right\} =$$

$$= (c_1a_1^6 + c_2a_2^2)\delta(h;u) - 3a_1^4c_1\frac{\partial^2\delta(h;u)}{\partial h^2} + 3a_1^2c_1\frac{\partial^4\delta(h;u)}{\partial h^4} - c_1\frac{\partial^6\delta(h;u)}{\partial h^6} - c_2\frac{\partial^2\delta(h;u)}{\partial u^2}.$$

Consequently, Green's function could be derived through

$$G(h;u) = (c_1a_1^6 + c_2a_2^2)C_0(h;u) - 3a_1^4c_1\frac{\partial^2 C_0(h;u)}{\partial h^2} + 3a_1^2c_1\frac{\partial^4 C_0(h;u)}{\partial h^4} +$$

$$-c_1 \frac{\partial^6 C_0(h;u)}{\partial h^6} - c_2 \frac{\partial^2 C_0(h;u)}{\partial u^2} \tag{5.4.3}$$

provided that $C_0(h;u)$ and its derivatives are available. It surely is challenging to derive $C_0(h;u)$ analytically from the Fourier transform of $C(h;u)$ in (5.4.1); but numerical representations of $C(h;u)$, as well as of the respective partial derivatives, are easily found. Figure 5.13 shows the function in space-time domain, by the Discrete Fourier Transform (DFT, only the 1st quadrant is shown), Both second numerical and fourth numerical derivatives, with respect to *h*, are shown in Figure 5.14 and Figure 15,

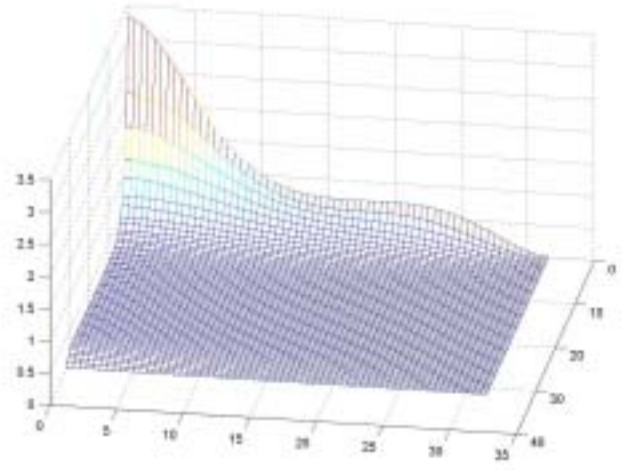


Figure 5.13: Spectral representation of the covariance function (5.4.1) by DFT

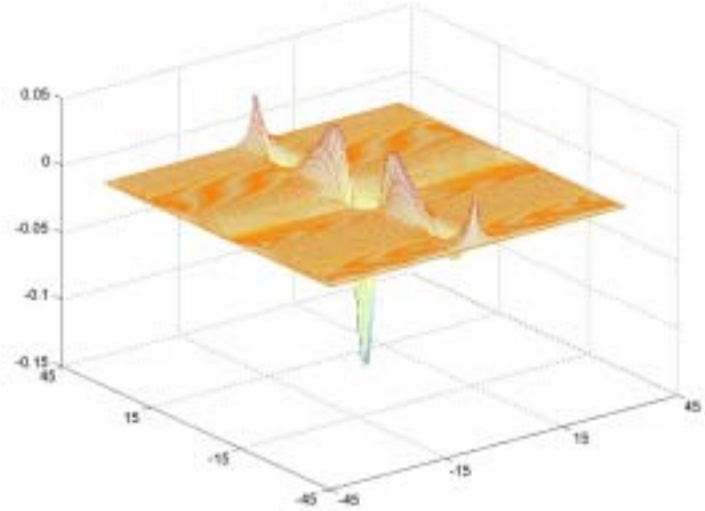


Figure 5.14: Numerically derived 2nd order derivative of $C_0(h;u)$

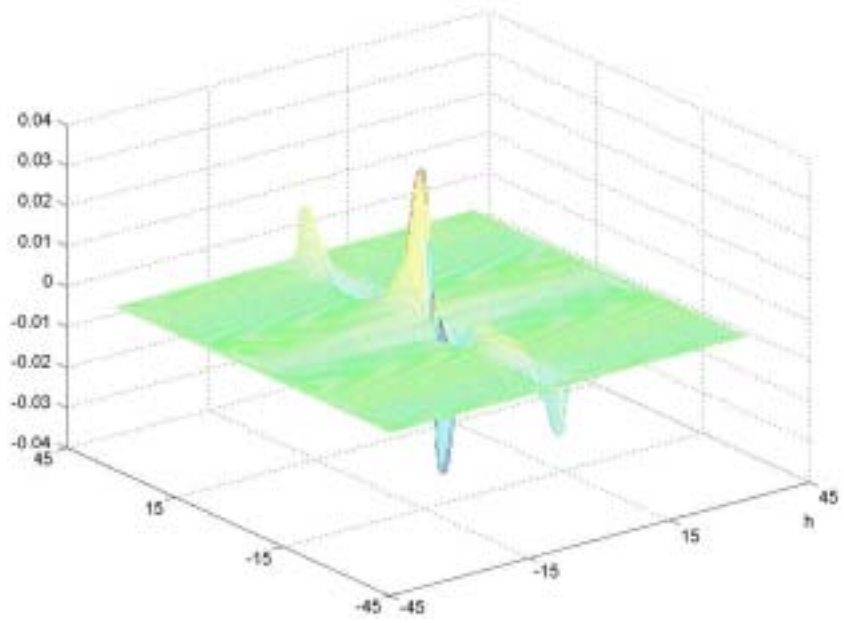


Figure 5.15: Numerically derived 4th order derivative of $C_0(h;u)$

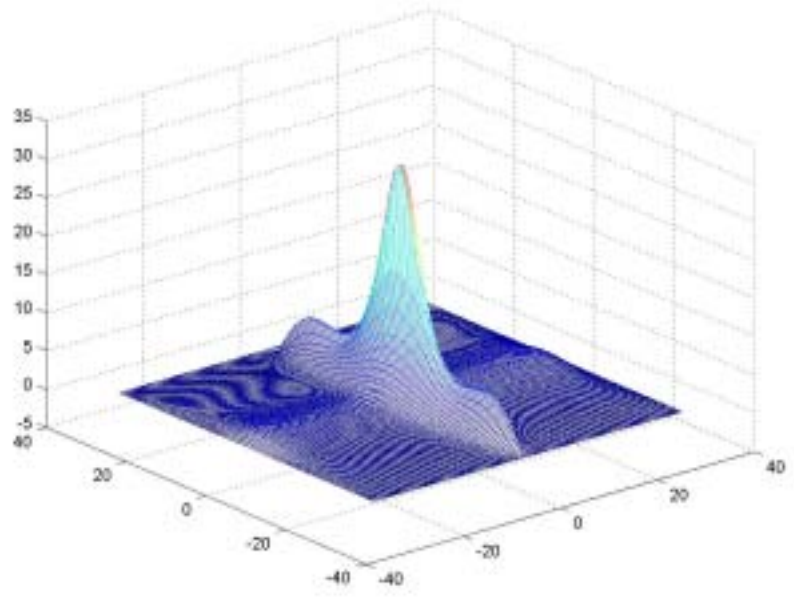


Figure 5.16: The Green's function as computed from equation (5.4.3)

respectively. Finally, the Green's function (5.4.3) is illustrated in Figure 5.16 (in non-normalized form).

5.5 Determining the Interpolator by Identifying the Covariance Function

In order to demonstrate the processing of data, a simulated random field is employed in this study. One basic principle to simulate a dataset with spatial or temporal correlation consists in the factorization of the related variance-covariance matrix. In this perspective, matrix factorization provides a way to isolate one of the random variables to be independent from the others, and by choosing the structure of lower or upper triangular matrix, all other random variables can be simulated sequentially, based on the first independent initial realization. The actual decomposition of a matrix, however, would cause a tremendous challenge should the size of the matrix increase considerably.

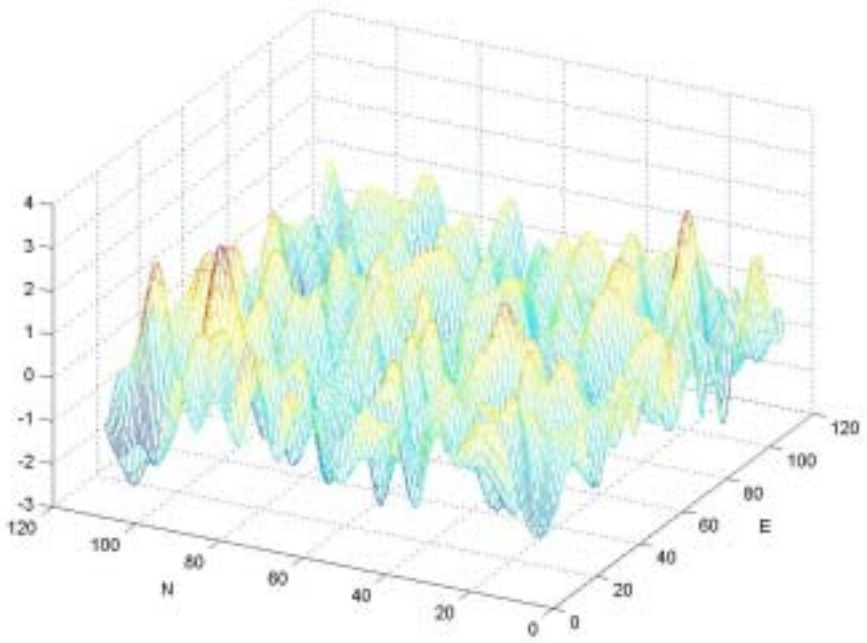


Figure 5.17: Simulated stationary random dataset for a given covariance function

Chan and Wood (1997) proposed an algorithm to simulate a Gaussian random field defined on a d-dimensional Euclidean cube $[0,1]^d$. In their algorithm, the eigenvalue decomposition is excuted by applying the transformation into the spectral domain. After the sequential simulation is carried out, the simulated data of the random field are available by backward transformation into the original domain. In addition, to take care of any edge effects, the variance-covariance matrix, which is known to be a Toeplitz matrix, has to be augmented in advance by periodical duplications and will so become a circulant matrix. Figure 5.17 shows the simulated data of a regular grid field with 128×128 sites

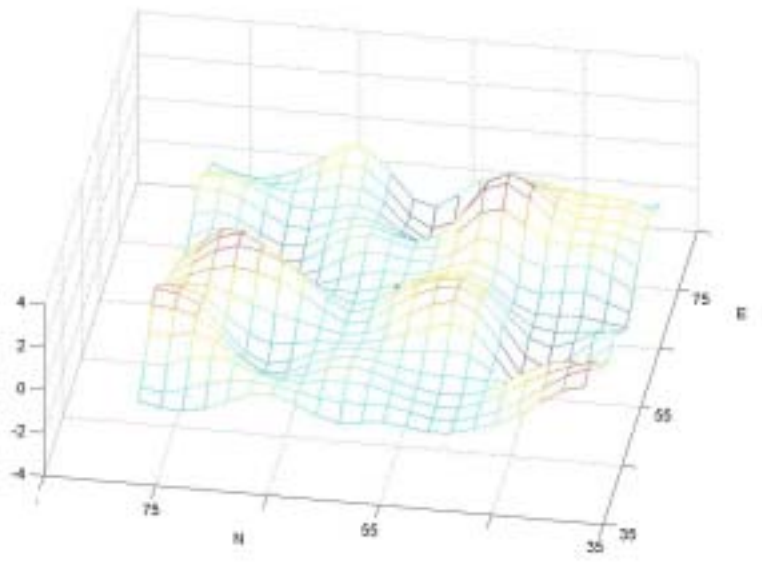


Figure 5.18: The close-up view near the benchmark point x_0

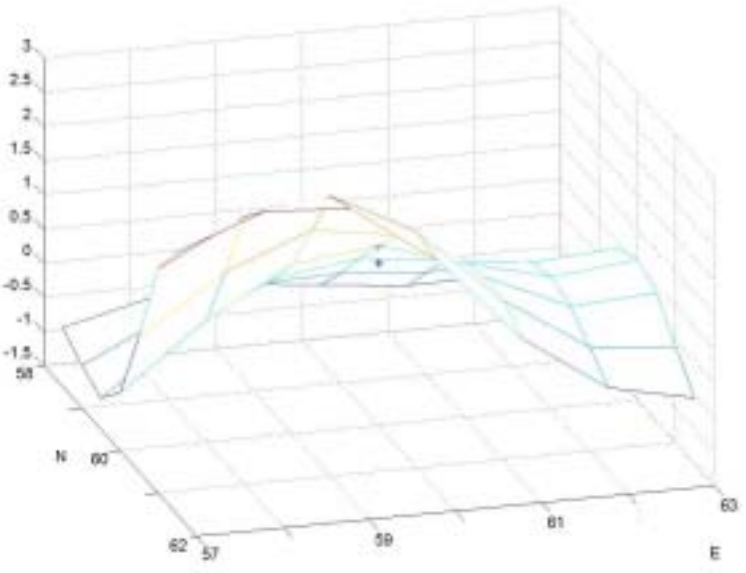


Figure 5.19: The benchmark point $Z(x_0)$ (marked by "+'), and the predicted value $\tilde{z}(x_0)$ (marked by "*") as derived from a 16×16 neighborhood.

After the identification of the covariance function and its parameters (2D-Gaussian with the only parameter σ^2), the Green's function can be calculated accordingly, following the approach discussed in Section 4.5. In order to demonstrate the

interpolation by Green's function, a data point $x_0 = (60,60)$, is chosen as a benchmark point, and its value $z(x_0)$ is marked by "+" in Figure 5.18, along with the nearby landscape adopted from Figure 5.17. Meanwhile, the predicted signal $\tilde{z}(x_0)$ is calculated using the data from the 16×16 neighborhood only and marked by "*" in Figure 5.19, which provides a better close-up view around site x_0 and shows the remaining discrepancy between $\tilde{z}(x_0)$ and $Z(x_0)$.

CHAPTER 6

APPLICATIONS IN GEODETIC SCIENCE

We have introduced the interpolation approach to deal with data sets of reasonably high sampling rate in Chapter 4 and demonstrated the numerical handling in Chapter 5. However, there may still exist a need to further enhance the sampling situation. Insofar as, when two datasets are to be analyzed jointly, they may often be required to have identical sampling configurations (in space and time). If not, then at least one of the datasets has to be later synchronized with the other one. This would be one of the situations in which to employ the interpolation technique.

Another occasion to apply such an interpolation technique is given in the case where some filtering is demanded prior to the follow-up data processing. Usually, a mask is empirically used to cover the data sites located within a neighborhood, before defining the functional relationship between the signal of the center site and all the other neighbor sites; this filter would sequentially move over the entire surveyed area and replace the raw signal at each site with a suitable combination from its neighbors. Following the principles developed here, one can could determine an optimal filter, in accordance with the properties prevalent within a specific dataset, by analyzing certain statistical measures.

6.1 Data Fusion of InSAR and LIDAR Data Sets

Slatton (2001) studied the issue of fusing InSAR and LIDAR data. Both datasets are topographic measurements collected by active sensors; the background principles of InSAR and LIDAR can be found in recent papers or textbooks, such as Hassen (2001). Generally, InSAR provides a broad coverage while LIDAR has the higher sampling rate compared with InSAR (along the track). In order to integrate the two datasets, the InSAR data, although in a quite fine grid format, ought to be further densified in order not to have to downgrade LIDAR's resolution (sampling rate) so that the two datasets can match in terms of sampling configuraton. In Slatton (2001), this task is performed by Kalman filtering where the conventional epoch index t is used to denote the layers of different resolutions, and one dataset is considered to be collected in the vector y, awaiting to be fused with the other dataset x at the resolution layer t:

$$\underset{n(t)\times 1}{y(t)} = A \cdot \underset{n(t)\times 1}{x(t)} + e(t), \qquad e(t) \sim (0, \Sigma_e), \tag{6.1.1}$$

$$\underset{n(t)\times 1}{x(t)} = \Phi \cdot \underset{n(t-1)\times 1}{x(t-1)} + w(t), \qquad w(t) \sim (0, \Sigma_w), \qquad C\{e(t), w(t)\} = 0, \tag{6.1.2}$$

where $n(t)$ and $n(t\text{-}1)$ denote the data volume at resolution layers t and t-1, respectively. Slatton (2001) sets up a coarse-to-fine and a fine-to-coarse pyramid scheme so that, in the state update equation (6.1.2), the fixed-size neighborhoods in layer t-1 would generate new data in layer t, which means each row in matrix Φ represents an interpolator drafting data from one neighborhood as illustrated in Figure 6.1.

Definitely, the various Kriging techniques will be among the most legitimate approaches to construct the matrix Φ in the state update equation, instead of the so-called empirical methods. Moreover, the update error $w(t)$ can be interpreted as prediction error if Kriging models are applied. In consideration of the high resolution of either the InSAR or LIDAR data set, the inversion approach proposed in this thesis will be of great help in the fusion process.

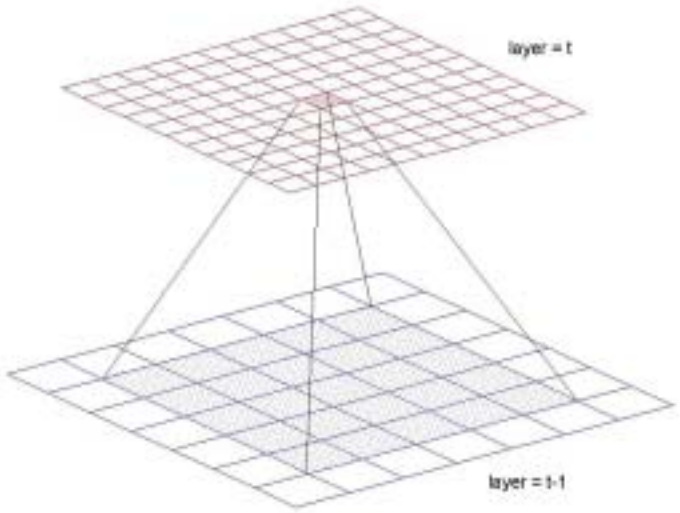


Figure 6.1: Data sets of different resolution (layer t and layer t-1). The pixel 0 in the layer t is interpolated by pixel 1, 2, 3, and 4 in the layer t-1.

To serve as a demonstration of the data fusion process, an InSAR and a LIDAR dataset, simulated by the random field simulation algorithm of Chan and Wood (1997), is adopted in our study. Two different situations, with assumed trend and with unknown trend, are discussed individually. Figure 6.2 shows the patch of a surveyed field where the InSAR terrain data have been generated in a regular grid format. Meanwhile, Figure 6.3 shows the result after the trend, treated as assumed or provided as information, has been removed, leading to a stationary zero-mean random field. This trend removal usually happens during the calibration phase, in which time the data collection is carried

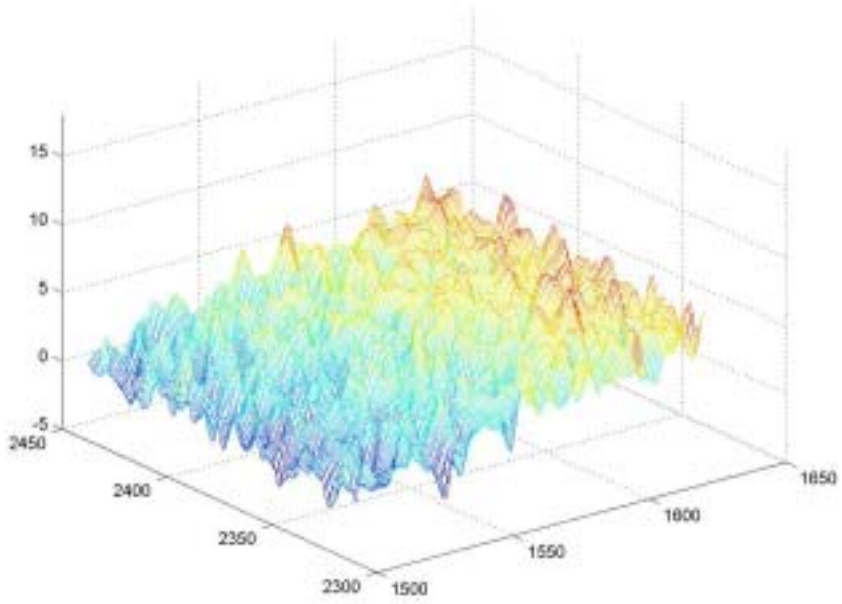


Figure 6.2: The surveyed field scanned by InSAR (F1_a)

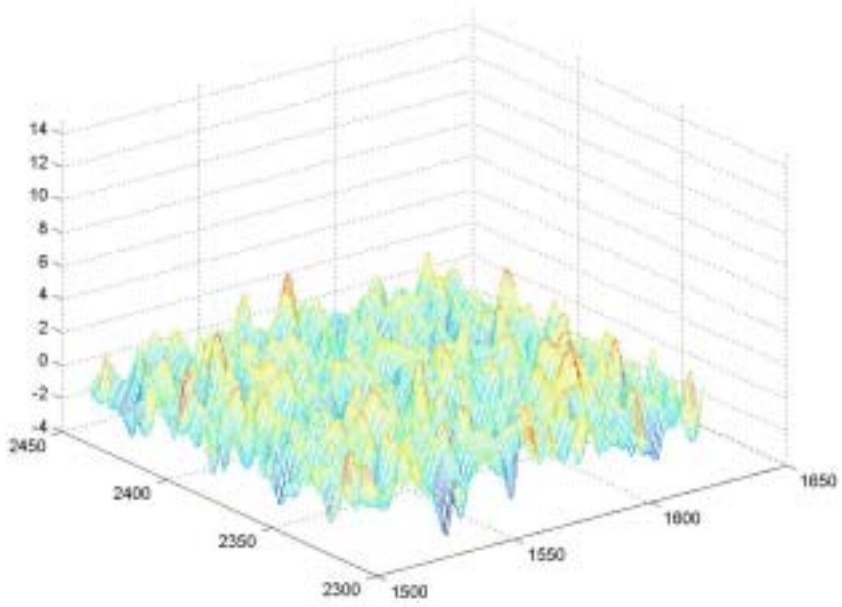


Figure 6.3: InSAR Dataset after the removal of trend (zero mean random field)

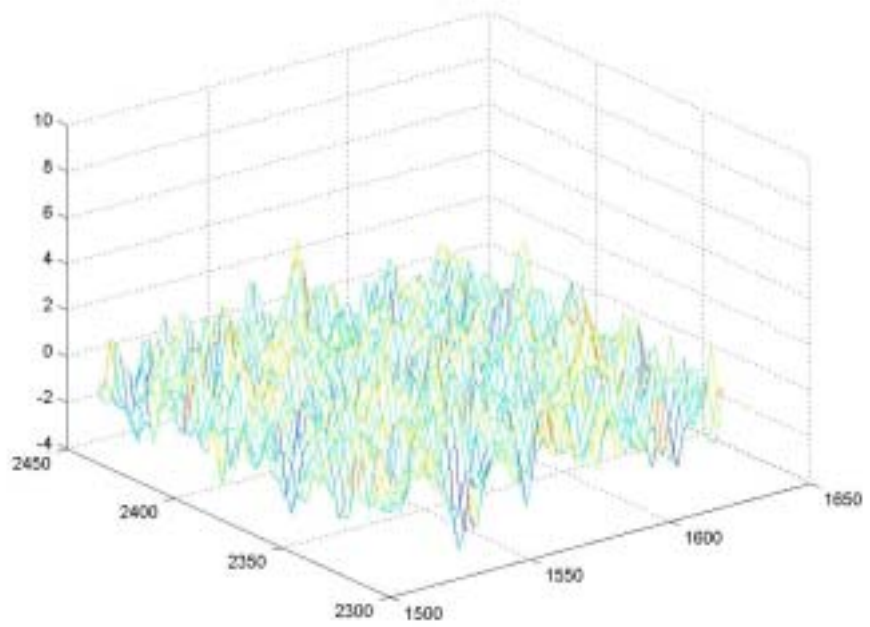


Figure 6.4 InSAR dataset at resolution t=2

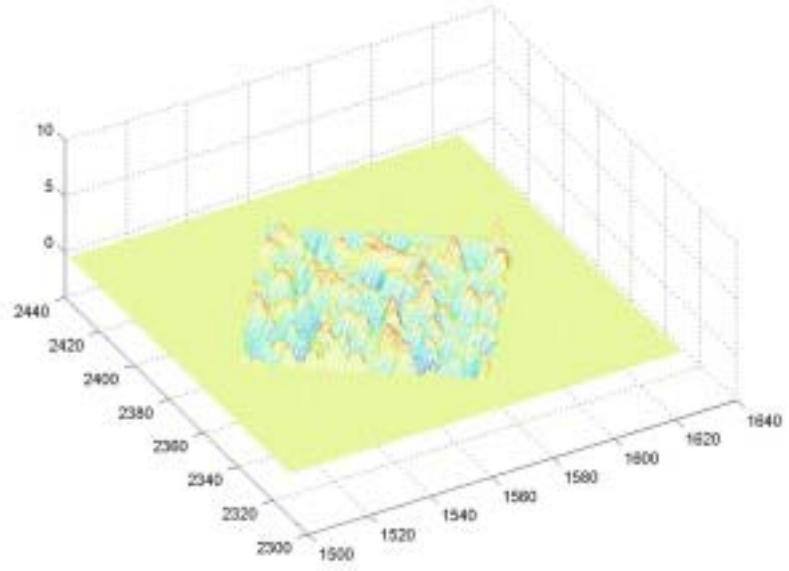


Figure 6.5 LIDAR dataset at resolution t=3 (F2_a)

out on a field where the "ground truth" has been well studied, so that the desired parameters can safely be determined, either for bias correction or stochastic modeling. According to the algorithm of Slatton (2001), the dataset is generated at different

resolutions hierarchically; here, the dataset of resolution t=2, with unit length (20 meter) in each grid, is designated as F1_a and illustrated in Figure 6.4.

Like the simulated InSAR dataset, the de-trended LIDAR dataset is designated as F2_a and shown in Figure 6.5. The F2_a file has been generated at the resolution t=3 (half the sampling interval to the previous layer t=2); due to different sampling characteristics however, the alignment of LIDAR data patch is oblique relative to the format of the F1_a file. The sampling schemes of both datasets are plotted in Figure 6.6, where the InSAR sites are marked by blue crosses, and the LIDAR sites by the red ones. A close-up look at the sampling sites of both datasets is provided in Figure 6.7.

The determination of the spatial covariance model usually relies on the information stemming from the sample covariance. By equation (2.1.4), the sample covariance of the F1_a file is computed in pairs, and the resulting scatter plot, is shown in Figure 6.8. The sample covariance data are then divided into groups, according to their ranges, to fit into the selected parametric model as seen in Figure 6.9. Since isotropy is not assumed in this case, sample covariance in various directions have to be analyzed individually. For the F1_a file, sample covariance in three sectors of the first quadrant were analyzed with consideration to generate even functions. The compound estimated covariance function is shown in Figure 6.10.

By the convolution theorem introduced in Section 4.6 and following the examples

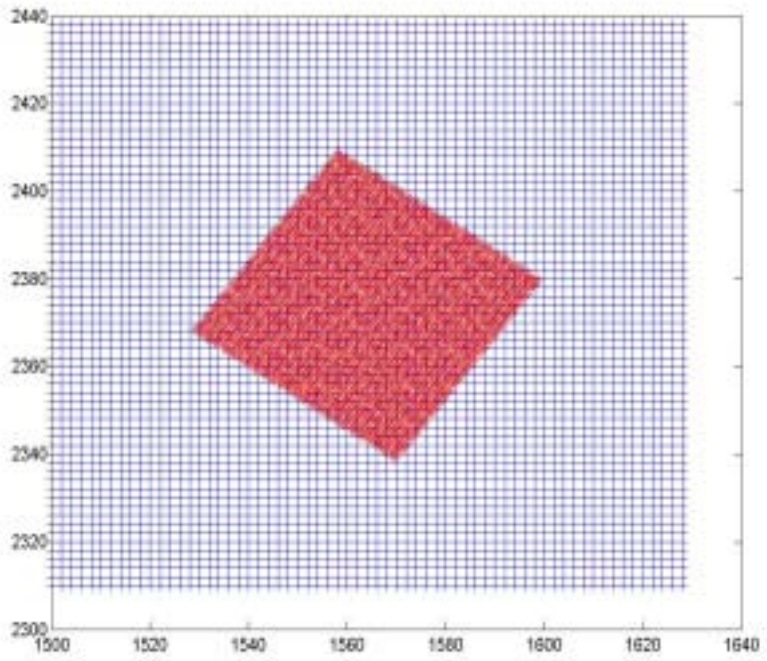


Figure 6.6: The data allocation sites of InSAR(blue) and LIDAR(red)

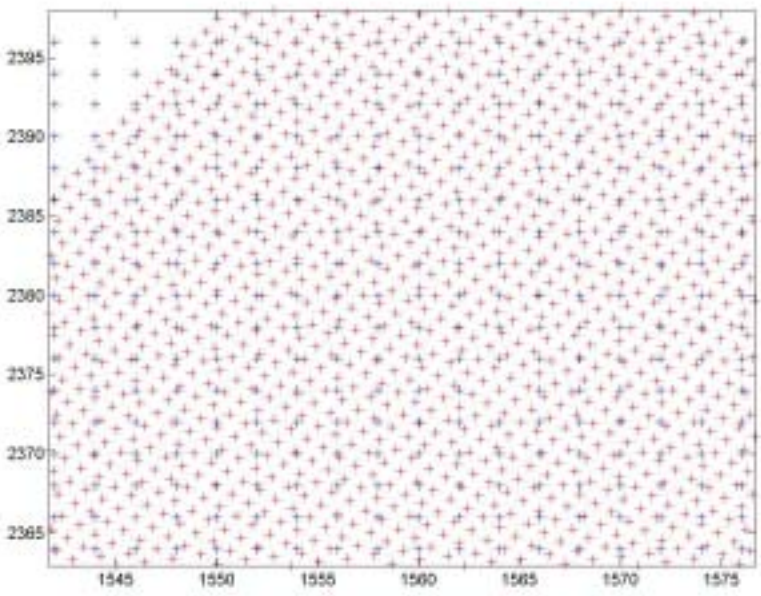


Figure 6.7: The data sites of InSAR and LIDAR in close-up view

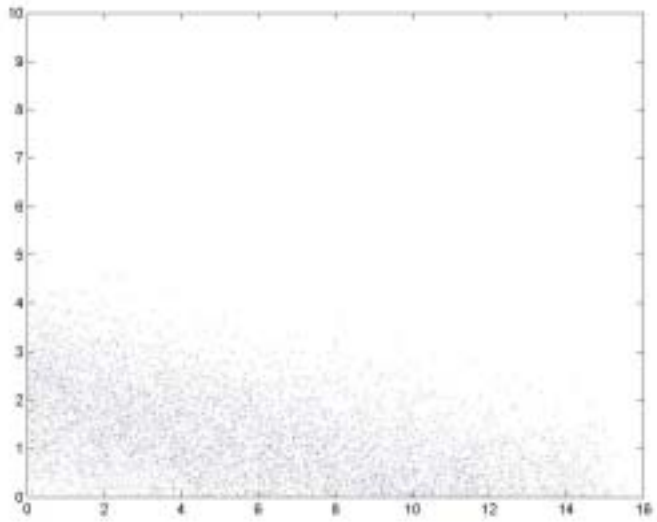


Figure 6.8: The scatter plot of the pairwise sample covariances (InSAR data)

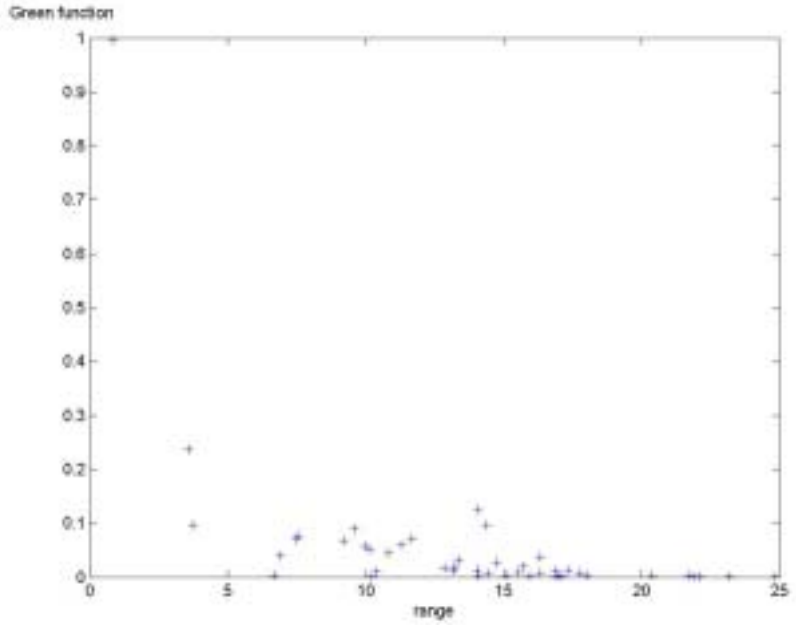


Figure 6.9: Discrete empirical covariance values (InSAR data)

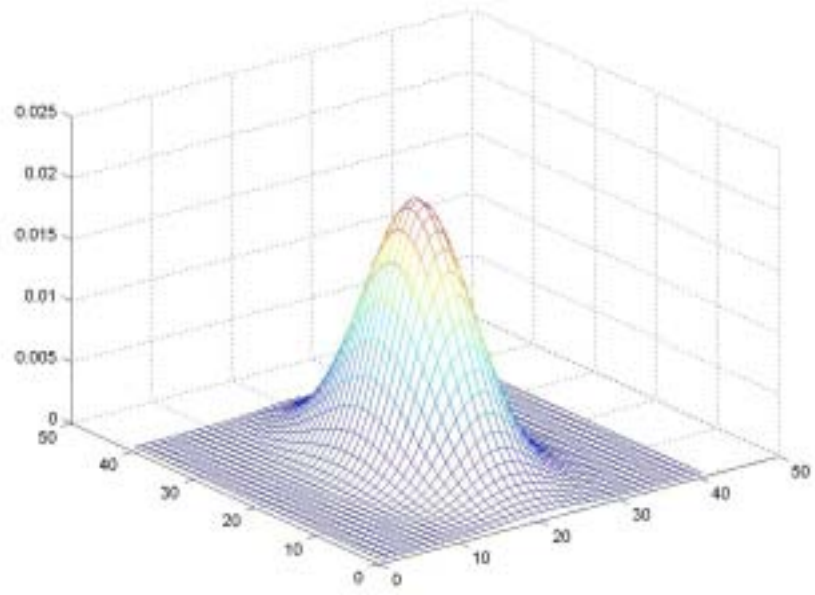


Figure 6.10: The compound estimated covariance function in a 40×40 mask matrix

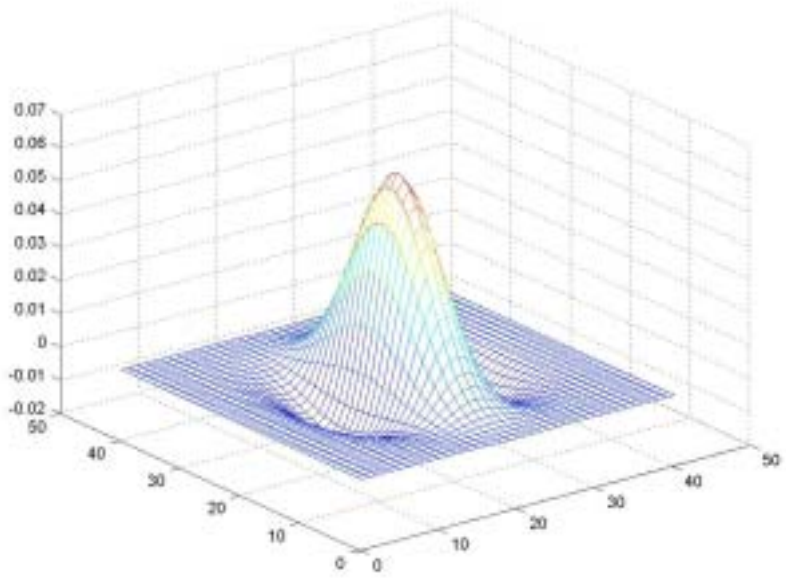


Figure 6.11: Calculated Green's function in a 40×40 mask matrix

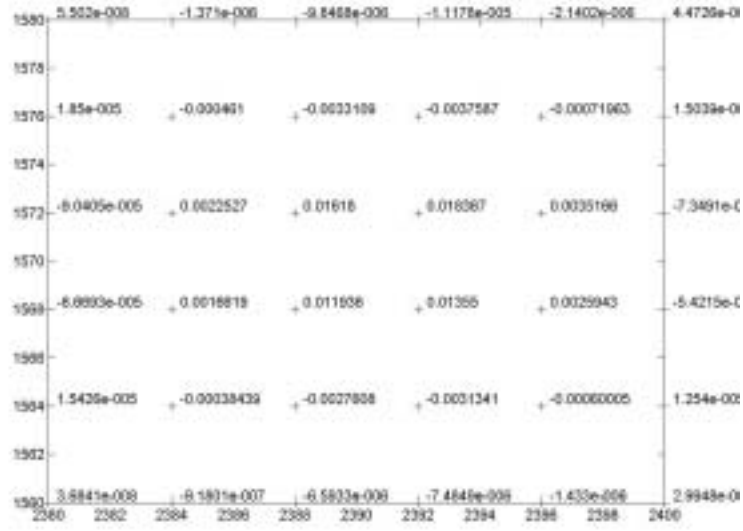


Figure 6.12: Coefficients of the interpolation mask at (2391,1573) for InSAR

in Section 5.3 and 5.4, the Green's function is now calculated with respect to the estimated covariance function, and plotted in Figure 6.11. with respect to a new location where the signal is to be interpolated, This Green's function is discretized by a suitable mask, so providing the elements for Φ in (6.1.2). Figure 6.12 shows such a mask for interpolation at the location (2391, 1573). Based on the Green's function, the InSAR dataset is transfered to a new dataset R1_a, which is entirely in synch with the dataset F2_a. The InSAR and LIDAR data are thereby fused together using the model (6.1.1-2),

and the result is shown in Figure 6.13. Incidentally, compared with the "true" terrain from the simulation, the residule of interpolation in InSAR dataset is plotted in Figure 6.14. The root mean squared prediction error is ±0.13 at its minimum level. The mean squared prediction error within a regular grid is plotted in Figure 6.15.

As indicated by Slatton (2001), the hierarchical transformation of datasets can as well be carried out the other way around. Namely, the dataset of higher resolution, F2_a, is downgraded into rougher resolution (t=2) and fused with F1_a by the same approach; only this time, the transform filter entails the use of the Green's function associated with the covariance function of the F2_a file. Figure 6.16 shows the dataset generalized from F2_a by the backward transform filter, and Figure 6.17 shows the final fusion result with F1_a at the t=2 resolution level.

Following the first experiment where datasets are assumed with given mean functions, the *second experiment* demonstrates a different situation where the mean functions are unknown. Figure 6.18 illustrates the landscape of the InSAR dataset F1_b, which represents the InSAR data collected from the second surveyed field. Like the dataset F1_a in the previous experiment, the dataset F1_b is of the same resolution t=2. On the other hand, Figure 6.19 shows the LIDAR data in file F2_b that were collected at a higher resolution (t=3), but over a smaller area inside the InSAR surveying ground. Figure 6.20 shows the configuration of two data patches, where the blue crosses mark the sites of the F1_b file and red crosses of the F2_b file.

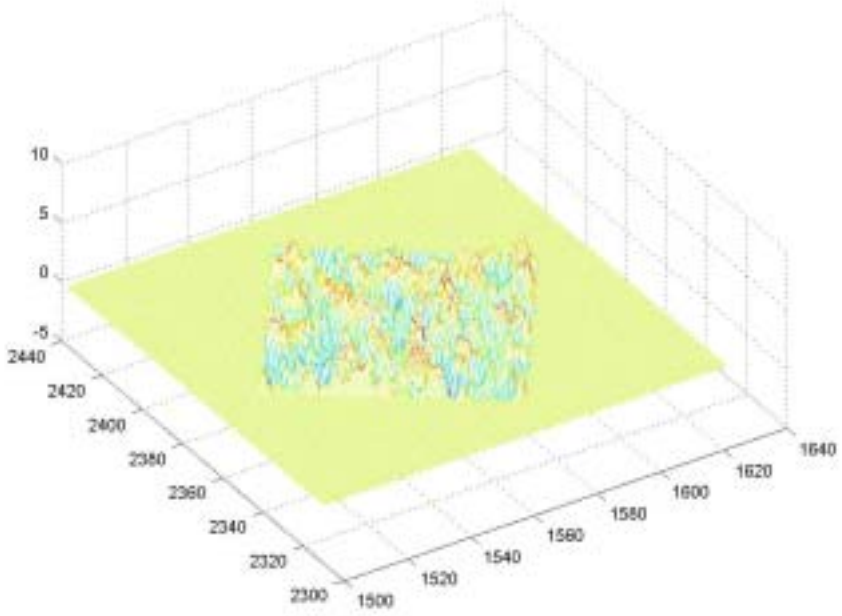


Figure 6.13: The fused InSAR and LIDAR datasets at resolution t=3

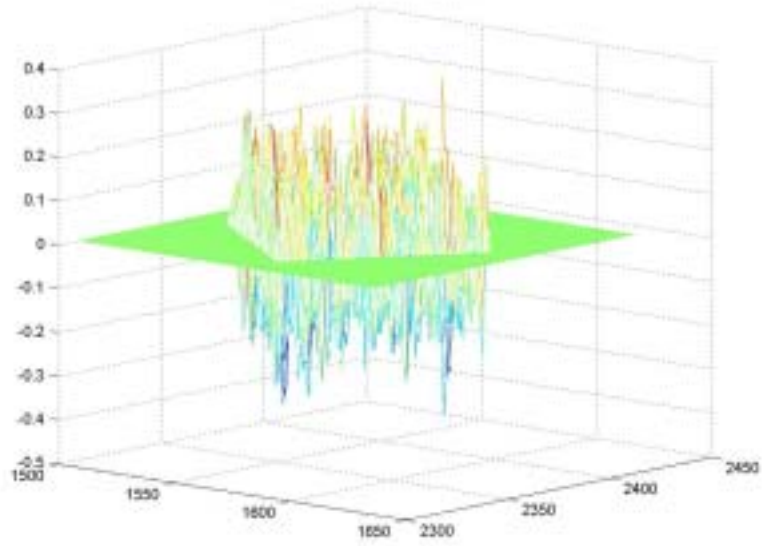


Figure 6.14: The residual of interpolation for InSAR dataset

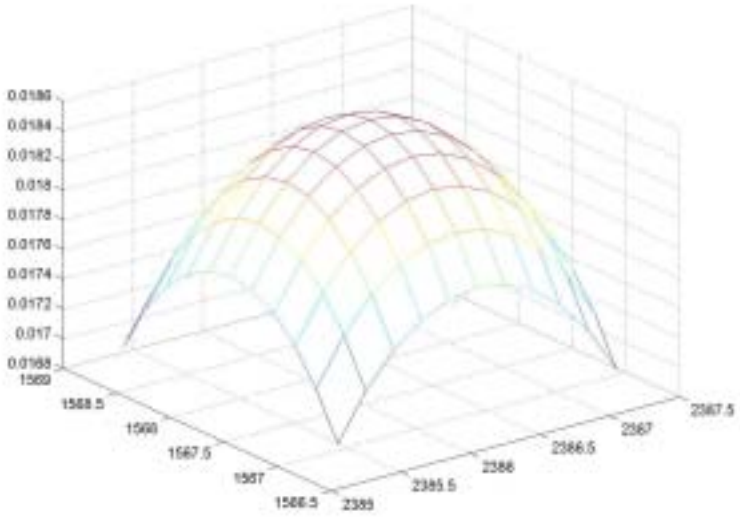


Figure 6.15: The plot of mean squared prediction error within a region of grid (in squared units)

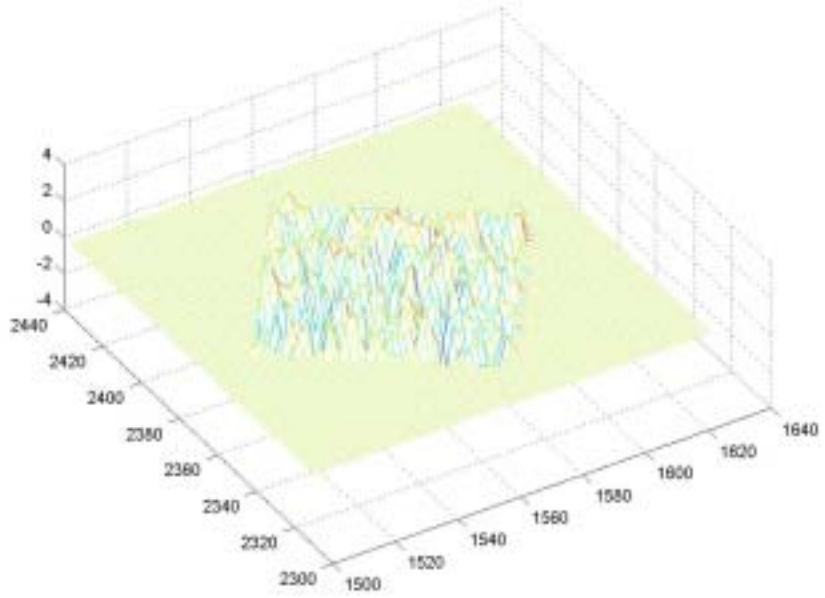


Figure 6.16: LIDAR dataset after downgrading its resolution to t=2

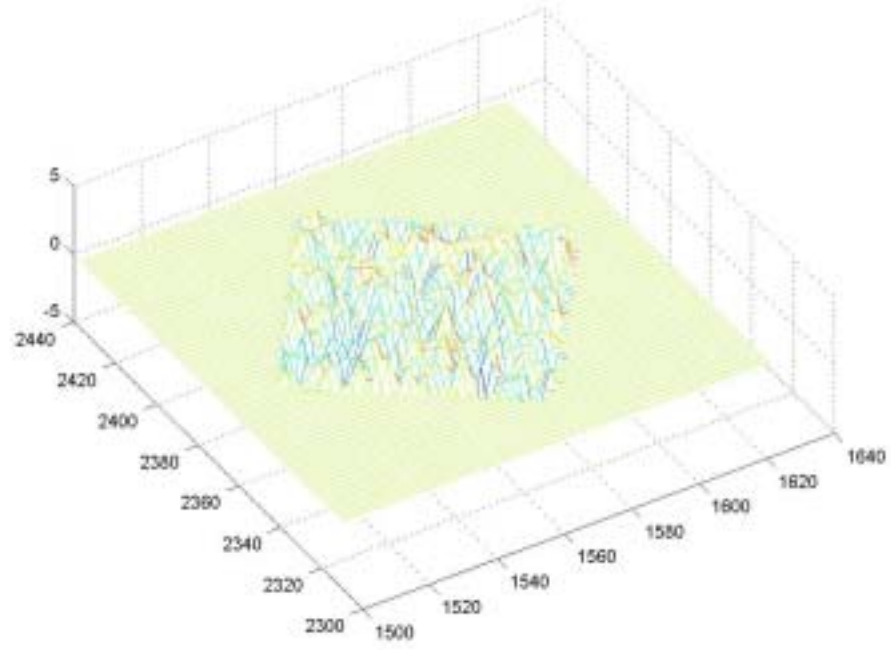


Figure 6.17: The fusion result of InSAR and LIDAR data at resolution t=2

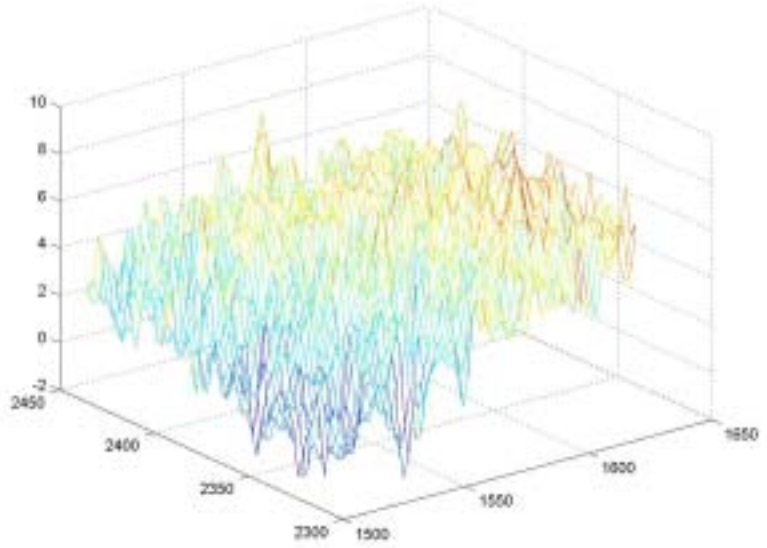


Figure 6.18: InSAR dataset with unknown mean (F1_b)

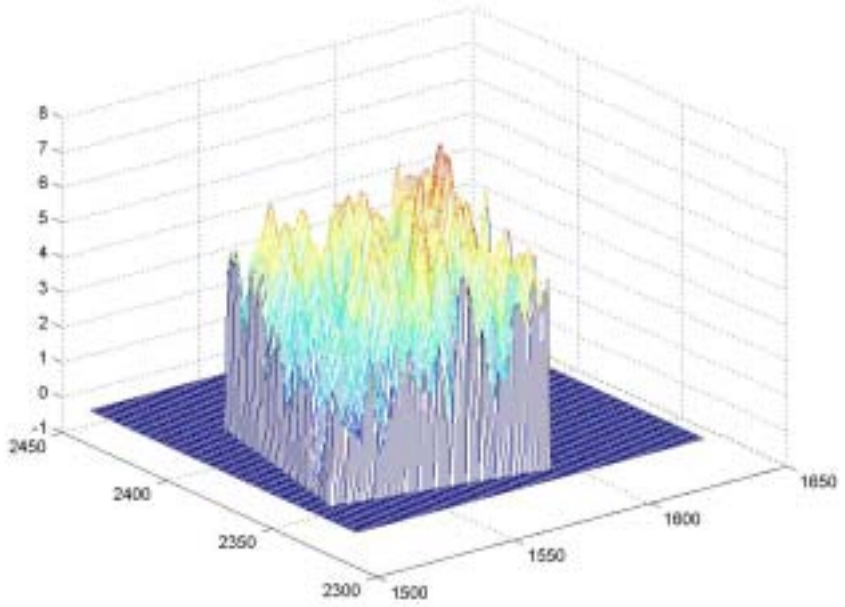


Figure 6.19: LIDAR dataset with unknown mean (F2_b).

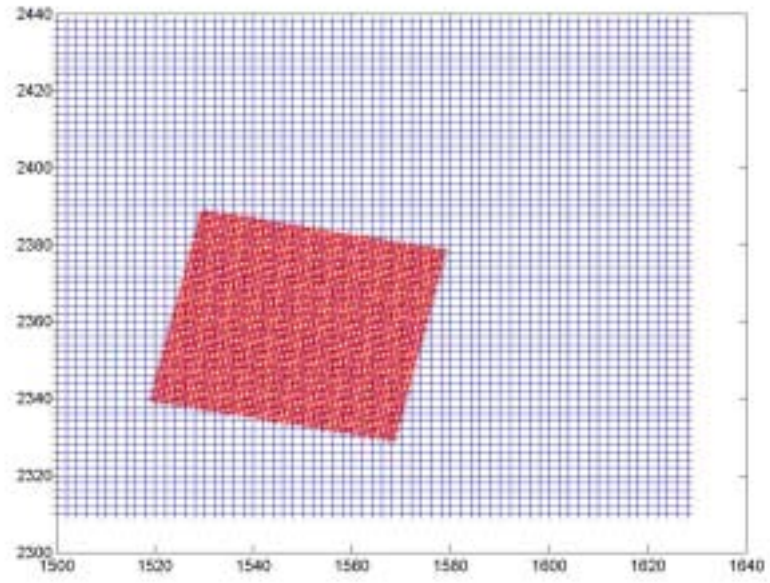


Figure 6.20: The overlap of InSAR(blue) and LIDAR(red)

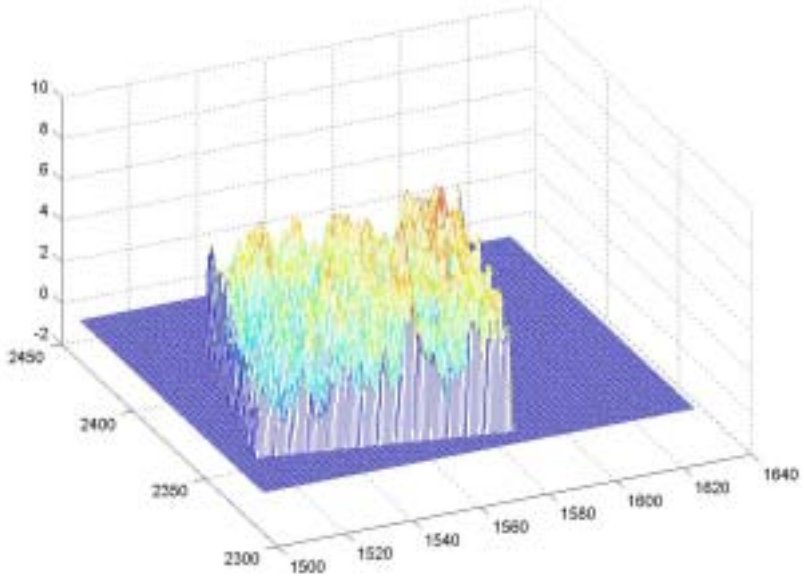


Figure 6.21: Interpolated InSAR data (from t=2 to t=3)

The covariance model of dataset F1_b is assumed to be given, and identical to that of dataset F1_a, which can be considered as stemming from the calibration site that serves to determine the stochastic model for all the following datasets. The Green's function of dataset F1_b is calculated through the Universal Kriging approach introduced in equation (2.4.12). Figure 6.21 shows the transformed InSAR dataset from resolution

level t=2 to t=3, it is designated as R1_b. Now the datasets R1_b and F2_b have a common registration for every data point, and can be combined to form a homogeneous dataset through the Kalman filtering data fusion approach; see Figure 6.22. The estimated mean function of dataset F1_b is illustrated in Figure 6.23.

Tthe backward transformation from the dataset F2_b (t=3) into the resolution level of the dataset F1_b (t=2) is shown in Figure 6.24, and the fusion result at the coarser level t=2 is shown in Figure 6.25.

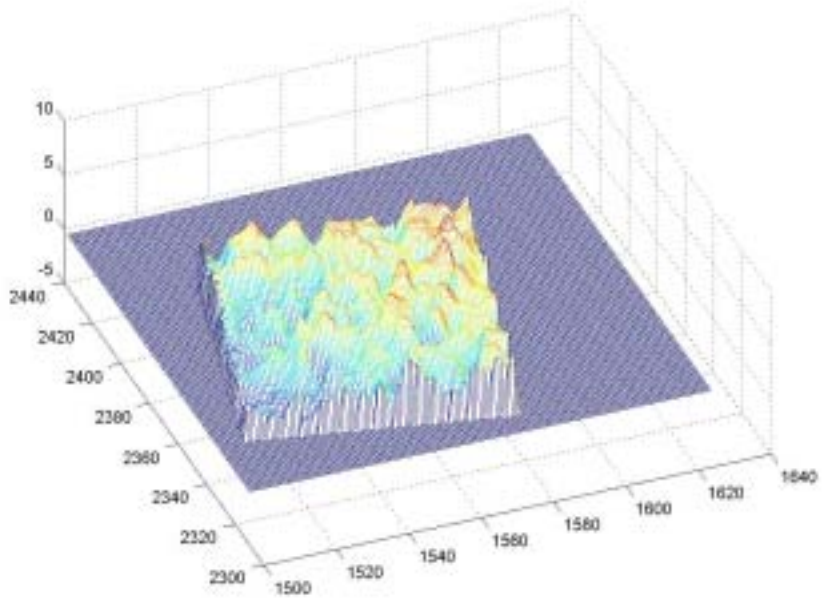


Figure 6.22: The fusion result of InSAR and LIDAR datasets at resolution t=3

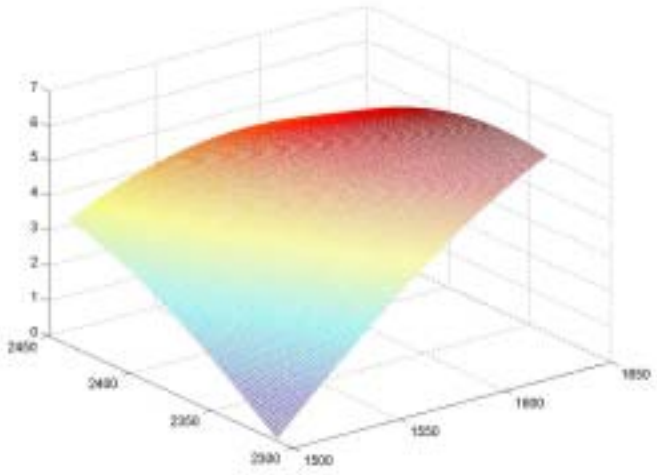


Figure 6.23: The estimated mean function of the InSAR dataset

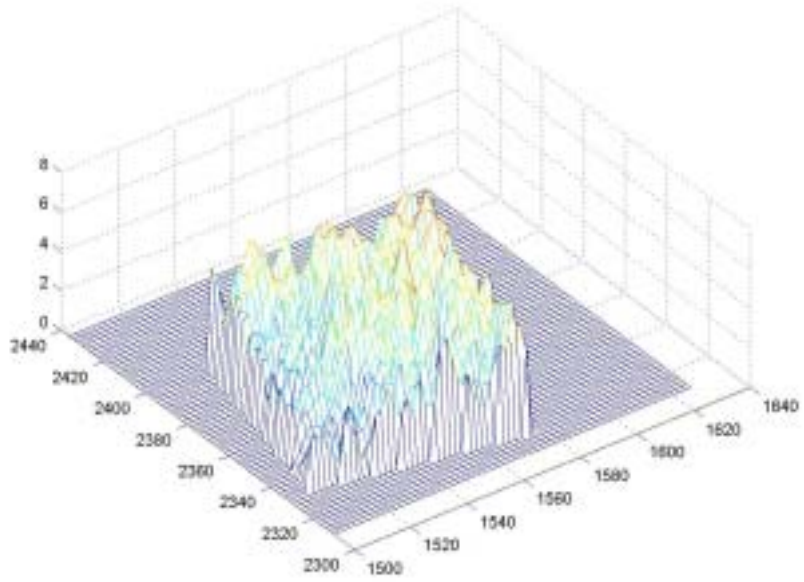


Figure 6.24: LIDAR dataset with downgraded resolution t=2.

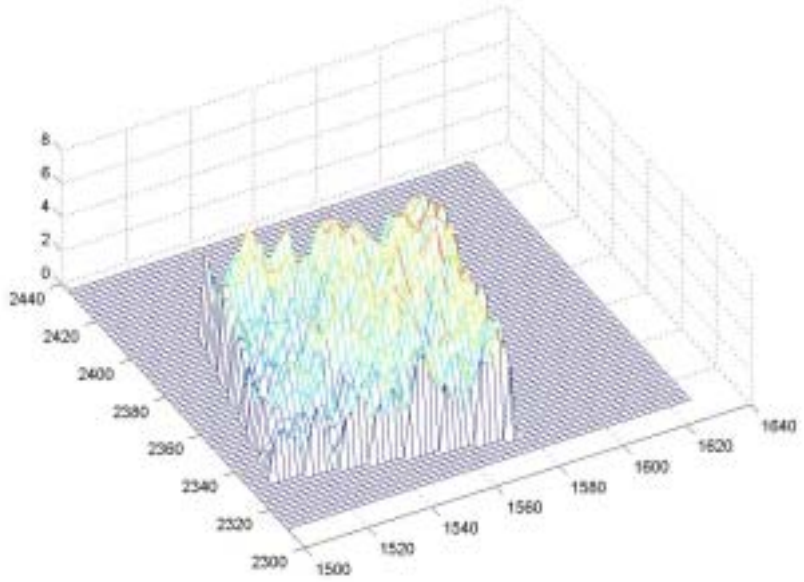


Figure 6.25: Fusion result of the two datasets at resolution t=2

6.2 Spatial-Temporal Interpolation for Terrain Profiles

The spatial-temporal analysis is a very attractive issue in geophysical studies. In order to demonstrate the interpolation of spatial-temporal datasets, we assume having

constantly scanned the terrain along a fixed profile-line at different time epoch. Dataset ST_1 simulates the monthly scanned terrain profiles in which the one-dimensional elevation data are recorded successively according to their scanning time. Figure 6.26 displays portion of the dataset in which the terrain profile scanned at time epochs 1 to 5 are plotted from top to bottom and the horizontal axis stands for the locations along the profile-line. The one-dimensional example can help us visually recognize the correlation existing between time epochs with respect to locations along the profile-line; we can see some terrain features gradually change from one profile to the next profile. Now the task is: could we generate a new elevation profile for the time between the two scanning epochs 3 and 4 (say)? In this dataset, the stationarity assumption holds, and the constant mean is given by μ=0.

Conceptually, the dataset ST_1 is a realization taken from a two-dimensional random field; every data point in the ST_1 file is denoted by $z(h;u) \in \mathbb{R}^1 \oplus \mathbb{R}^1$. However, since the indices involve both spatial and temporal dimensions, the underlying random process must assume a spatial-temporal covariance model. Figure 6.27 shows the estimated covariance function of type (5.2.1) for the dataset ST_1.

The associated Green's function can be determined after the covariance function has been estimated from the data. As a consequence, the supposed terrain profile at time epoch 3.5 is interpolated by the so specified interpolating mask as seen in Figure 6.28.

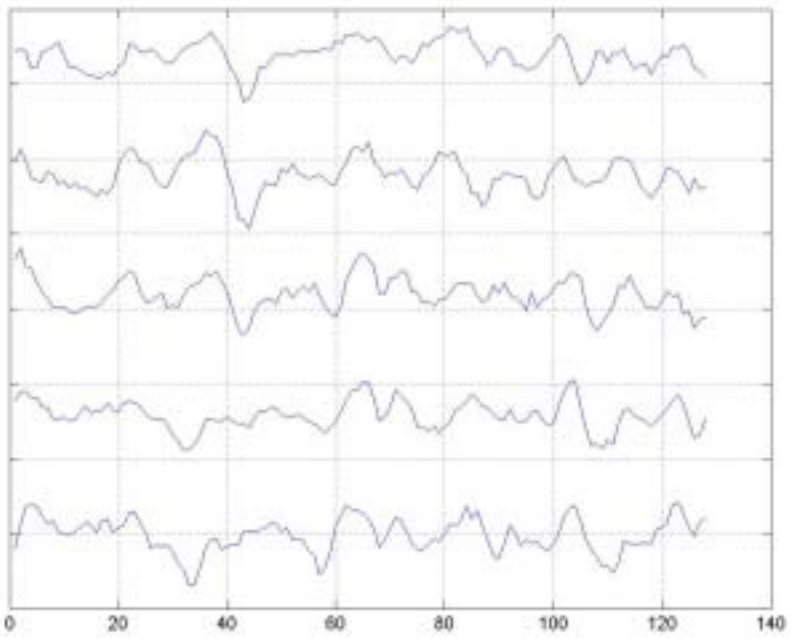


Figure 6.26: Terrain profiles from scannings at time epoch 1 to 5 (bottom to top)

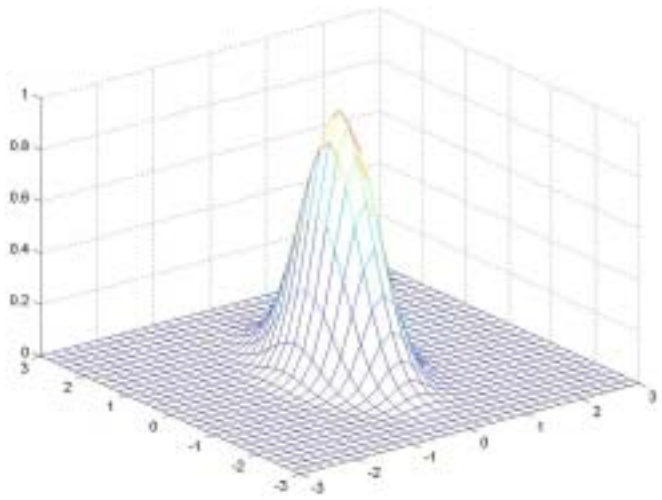


Figure 6.27: The estimated spatial-temporal covariance function

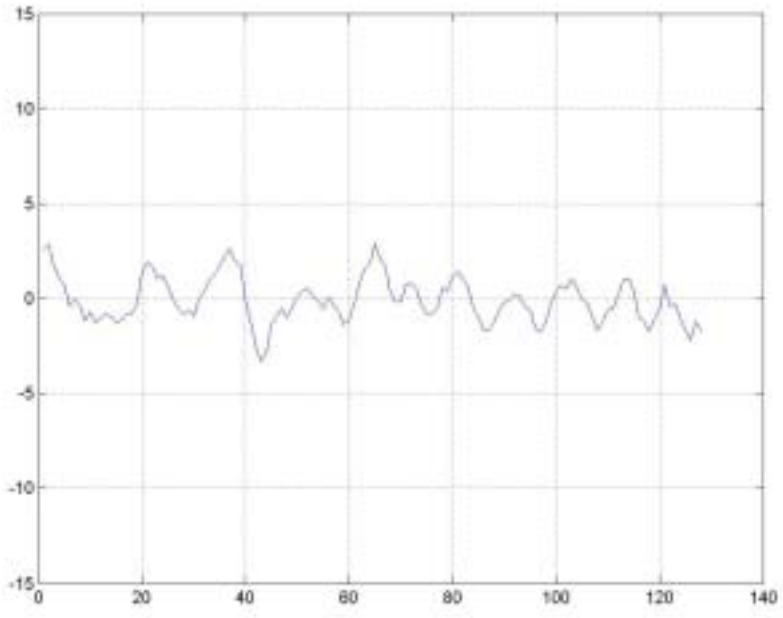


Figure 6.28: Interpolated terrain profile at time epoch 3.5

6.3 Filter Selection for de-Noising and Data Generalization: An Outlook

As the technology evolves, surveying instruments often collect massive amounts of data with a sampling rate higher than necessary; sometimes data overflows cause problems in the following analysis. Meanwhile, since every data set more or less carries some sort of unwanted random noise, the redundancy from the oversampling provides a possibility to enhance the signal.

The GPS observations used for altimeter calibration may serve as one example (Shum et al., 2003). The GPS antenna set-up at Marblehead had collected data at the sampling rate of 1Hz; the dataset was then resampled to a sparser sampling rate of 1/3600 Hz by averaging (equally weighted). Apparently the average with equal weights cannot replace the optimal filter for data generalization unless the temporal correlation is zero among successive sampling epochs.

The Kriging interpolator would be an appropriate choice in order to model the averaging filter, if spatial-temporal randomness is considered to be one of the factors governing the outcome. As indicated previously in this thesis, to fill-in the weight coefficients by the Green's function (or to compute Kriging coefficient in the discrete sense), one needs to either analyze the sample covariance function, or fit the data into a random field model. An optimal averaging filter could then be determined based on the theorems from Section 4.10 and 5.5.

CHAPTER 7

CONCLUSION AND OUTLOOK

In this thesis, a new perspective for the well known Kriging equation is introduced in order to keep the covariance model, as well as the unknown interpolator, in the continuous forms; thus some potential numerical problems in the manipulation of large matrices can be avoided. The continuous form of the Kriging equation in such a revised representation is solved either by an expansion in the space of so-called hyperdistributions or by the Fourier transform in the spectral space, both of which methods provide the desired interpolator in terms of a Green's function. Several well established covariance models serve as test examples whose corresponding interpolators (Green's functions) are calculated to demonstrate this approach. We learn that the ill-conditioned problem, in the case of a high sampling rate, should better be approximated in function space by some expansion techniques.

In addition to the advantage in its numerical computation, the representation of a continuous interpolator solution, in form of a Green's function, provides a possibility to connect the observations from a random field with the unknown parameters of a covariance model directly. Such an estimation may not only serve as an alternative in stochastic model calibration, but also to create an analysis tool for applications in signal processing. By using methodology introduced here, one can take many other tasks related to spatially and/or temporally referenced datasets.

As far as the spatial-temporal covariance models are concerned, the derivation of such new models is very much needed for various applications; but it proves to be not an easy task. New advanced covariance models of this kind, however, can most directly be derived by the convolution technique based on simpler models currently available. A number of guidelines are suggested here to assist with the construction of such new models.

More importantly, even when physical knowledge is not available for a dataset, the covariance function can still reflect some of the physical behavior that would otherwise be expressed as a boundary value problem. Such a perspective provides as an alternative approach with respect to physical modeling; the massive dataset may give hints towards yet concealed physical properties before the physical model can be developed.

Certainly, there are still open questions worth pursuing after the conclusion of the present research. The first issue is to generalize the above results to the spherical domain. So far, to simplify the presentation of the new techniques, all the functions have been

defined in Euclidean space, represented in the Cartesian coordinate system. The representation on the sphere, however, is oftentimes more suitable for applications in geodesy. The interpolator approximated by the Green's function may leave some hints for the construction of appropriate spherical basis functions.

Secondly, the estimator of a random field still deserves more investigation. Here, only a simple likehood-based estimator has been applied, which still leaves much to ask for in terms of the tight interconnection between likehood and covariance parameters. In the future, however, an estimator based on the least-squares approach may show some promising efficiency. Moreover, as far as the random field estimator is concerned, the model selection (for the covariance in this case) is always a challenging, but nonetheless attractive issue.

Finally, the application of this approach to more problems from geodetic science is surely a desirable task. It could both help the analysis of specific datasets in certain scientific projects, and further the efficiency of this approach.

REFERENCES

Arfken, G. (1985), *Mathematical Methods for Physicists*, 3rd ed., Academic Press: Orlando, FL.

Arsac, J. (1966), *Fourier Transforms and the Theory of Distributions*, Prentice-Hall: Englewood Cliffs, NJ.

Bachman, G., L. Narici, and E. Beckenstein (2000), *Fourier and Wavelet Analysis*, Springer: New York. etc.

Besag, J. E. (1974), Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society* B-**36**, 192-225.

Bochner, S. and W. Martin (1948), *Several Complex Variables*, Princeton University Press, Princeton, NJ.

Bochner, S. (1955), *Harmonic Analysis and the Theory of Probability*, University of California Press: Berkeley and Los Angeles/CA.

Borre, K. (2001), *Plane Networks and Their Applications*, Birkhäuser: Boston.

Bracewell, R. (1968), *The Fourier Transform*, McGraw-Hill, New York.

Bruce, K. and E. Westwig (1998), *Mathematical Physics: Applied Mathematics for Scientists and Engineers*, Wiley: New York.

Chan, G. and A. Wood (1997), An algorithm for simulating stationary Gaussian Random Fields, *Applied Statistics*, **46**, 171-181.

Chilès, J. and P. Delfiner (1999), *Geostatistics. Modeling Spatial Uncertainty*, Wiley-Interscience, New York.

Cressie, N. (1991), *Statistics for Spatial Data*, Wiley: New York.

Cressie, N. and H. C. Huang (1999), Classes of nonseparable, spatio-temporal stationary covariance functions, *Journal of the American Statistical Association*, **94**, 1330-1340.

Dirac, P. A. M. (1958), *Quantum Mechanics*, 4th edn, Oxford University Press: Oxford, U.K.

Fröberg, C. (1965), *Introduction to Numerical Analysis*, 2nd edn, Addison-Wesley: Cambridge, MI.

Garabedian, P. (1964), *Partial Differential Equations*, Wiley: New York.

Gel'fand, I. M. and G.E. Shilov (1964), *Generalized Functions. Volume 1: Properties and Operations*, Academic Press: New York.

Gneiting, T. (2002), Nonseparable, stationary covariance functions for space-time data, *J. of the Ameri. Statist. Assoc.*, **97**, 590-600.

Hammersley, J. and P. Clifford, (1971), Markov fields on finite graphs and lattices, Unpublished manuscript, Oxford University Oxford/U.K.

Hanssen, R. (2001), *Radar Interferometry: Data Interpretation and Error Analysis*, Kluwer Academic: Boston, etc.

Hohlfeld, R., J. King, W. Drueding and G. Sandri (1993), Solution of Convolution Integral Equations by the Method of Differential Inversion, *SIAM J. Applied Math*. **53**, 154-167.

Kreyszig, E. (1983), *Advanced Engineering Mathematics*, 5th edn, Wiley: New York.

Krumm, F. (1987), *Geodetic Networks in a Continuum: Inversion-free Adjustment and Construction of Criterion Matrices* (in German), Publ. of the German Geodetic Comm. C-334, Munich/Germany

Loan, C. van (1992), *Computational Frameworks for the Fast Fourier Transform*, SIAM., Philadelphia, USA.

Matérn, B. (1960), *Spatial Variation*, 2nd. edn, Springer Lecture Notes in Statistics No.36, New York, pp.289-298.

Oberhettinger, F. (1990), *Tables of Fourier Transforms and Fourier Transforms of Distributions*, Springer, New York etc.

Oliver, D. (1998), Calculation of the Inverse of the Covariance, *Mathematical Geology* **30**, 911-933.

Schaffrin, B. (2001), Equivalent Systems for Various Forms of Kriging, Including Least-Squares Collocation, *Z. für Vermessungswesen* **126** (2001), 87-94.

Schaffrin, B. (2003), A note on constrained total least-squares estimation, manuscript submitted to *Computat. Statistics & Data Analysis*, Oct. 2003.

Schaffrin, B. (2004), Some Generalized Equivalence Theorems for Least-Squares Adjustment, *Springer Series, IAG-Symp*. No. 127, Berlin etc., pp.107-112.

Schwarz, K. P., M.G. Sideris and R. Forsberg (1990), The Use of FFT Techniques in Physical Geodesy, *Geophys. J. Intl*. **100**, 485-514.

Shum, C. K., Y. Yi, K. Cheng, and C. Kuo (2003), Calibration of JASON-1 Altimeter over Lake Erie, *Marine Geodesy* **26**, 335-354.

Slatton, K. C. (2001), *Adaptive Multiscale Estimation for Fusing Image Data*, Ph.D. dissertation, Department of Electrical and Computer Engineering, The University of Texas at Austin/TX.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer: New York, etc.

Stein, M. L. (2003), Space-Time Covariance Functions, Technical Report No 4, Center for Integrating Statistical and Environmental Science, University of Chicago.

Sveshnikov, A. (1978), Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions, Dover: New York.

Werner, C. (1994), *Lidar Techniques for Remote Sensing*, Proceedings of Europto, No. 2310, Bellingham/WA.

Winkler, G. (1995), *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer: Berlin.